ADVANCING RNA CIRCUITRY ENGINEERING WITH GUIDANCE OF MATHEMATICAL AND COMPUTATIONAL MODELS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Yongxi Hu

December 2018

ADVANCING RNA CIRCUITRY ENGINEERING WITH GUIDANCE OF MATHEMATICAL AND COMPUTATIONAL MODELS

Yongxi Hu, Ph.D.

Cornell University 2018

The main goal of synthetic biology is to harness the power of biological genetic expression in order to perform various tasks, which could often have impact on many aspects of our lives ranging from health care, pharmaceutical, to renewable energy and environments. These tasks are often carried out through synthetic genetic circuits composed of genetic parts like promoters, RBS, terminators and inducible gene regulators. Numerous studies have expanded the toolbox for synthetic circuitry and established the guideline of circuitry assembly in the past decade. As the field advances, regulatory sRNA is emerging as a powerful tool in synthetic biology. These regulatory sRNAs are versatile and designable, more importantly, they offer a fast dynamic in genetic circuitry due to the fast production rates and degradation rates of RNA molecules. However, little work has been done to create a theoretical foundation for RNA circuit design. With the ultimate goal of gaining full control of RNA circuitry dynamics, this work focuses on building the theoretical foundation to understand and guide the design of RNA synthetic circuitry.

A prerequisite to building such foundation is to create a modeling framework that accurately describes the dynamics of RNA circuits. In the first part of this work, we build an effective model composed of ordinary differential equations to describe transcriptional RNA genetic circuitry, and validate the model using a three level cascade as a test case. We develop a sensitivity analysis based parameterization procedure that requires only a handful of simple experiments that can be performed in parallel using

rapid cell-free transcription-translation (TX-TL) reactions. This part of the work establishes a fundamental method to predict circuit dynamics, which allow us to build more complex systems. Next we expand the repertoire of synthetic gene networks built from these regulators by constructing a transcriptional negative autoregulation (NAR) network out of small RNAs (sRNAs). Using parameter sensitivity analysis, we design a simple set of experiments that allow us to accurately predict NAR circuit dynamic in TX-TL. We also transfer the successful network design into *Escherichia coli* and show that our sRNA transcriptional network is functional *in vivo*. In the third part of this work, we investigate into an interesting observation where the transfer functions of inducible systems are altered by small transcriptional activating RNAs (STARs). We combine theory, computational model and experimental results to uncover the underlying causes of this phenomenon. Based on which, we establish the design principle of transfer function manipulation and dynamic range amplification induced by STARs.

Together the work presented here establishes a theoretical foundation of RNA circuitry design. We anticipate that this foundation would support the development of new mathematical and computational models that provide insights and guidance for RNA circuitry design —which ultimately contributes to the advancement of synthetic circuitry engineering.

BIOGRAPHICAL SKETCH

Chelsea (Yongxi) Hu was born and raised in Liuzhou, Guangxi, China. She is the only child to her parents Nangeng —a software engineer, and Hong —an physician and neonatologist. Chelsea grew up very close to her grandmother, Lian, who lead a close and loving extended family, which gave Chelsea a preciously joyous childhood. Chelsea's childhood dream was to become a scientist ever since her mother took her to the Beijing Planetarium when she was five-years old. However, her dream was later deemed unrealistic and naive by her middle school science teachers, as she constantly delivered disappointing test scores on science subjects. She was eventually convinced that she wasn't smart enough to do science. After high school, Chelsea went to Sichuan International Studies University in Chongqing to study English-Chinese translation.

Chelsea didn't like being a translator. After three unfulfilling years in college, she finally decided to pursue her true passion. She dropped out of college and moved to Los Angeles, where she started at a community college as a freshman to study science. Two years later, she transferred into UCLA as a junior with straight A's to study Chemical Engineering. She didn't really know what Chemical Engineering was, she only picked this major because it was rumored "the most difficult major at UCLA". As she struggled through all the math-heavy chemical engineering major classes, she developed an analytical understanding and a profound passion in complex systems and their dynamics. At UCLA, Chelsea also worked in professor James Liao's lab as an undergraduate student. She was mentored by a then Liao lab postdoc Dr. Christina Agapakis, who introduced her to synthetic biology and inspired her to go to graduate school to pursue her Ph.D.

After graduating from UCLA, Chelsea moved to Ithaca as a Cornell Ph.D. student in Chemical engineering. She considers herself incredibly lucky that she was able to do exactly what she proposed to do in her graduate school application —merging synthetic

biology and systems biology —as she was co-advised by two brilliant and supportive professors Julius Lucks and Jefferey Varner. Chelsea loved Ithaca. She made most of her friends at the Cornell Big Red Barn, where she worked as a student manager. She enjoyed running and hiking in Ithaca with her running buddies from the ChemE department.

Chelsea later moved to Northwestern University in Evanston in her fourth year with the Lucks lab. Next, her journey will take her back to LA for a postdoc position at Caltech to study control theory under Dr. Richard Murray. It has been 10 years since Chelsea left home to pursue her dream. She truly misses her family and friends back home, but she has never regretted a thing about her journey.

In the loving memory of

My grandmother Lian, whom I will be forever indebted to.

And Jenny —you will always be a part of me till the end of time.

ACKNOWLEDGEMENTS

To Julius, I'm incredibly lucky to have you as my advisor, I can't thank you enough for your support and guidance over the years. Despite all the funding pressure on you, you offered me total freedom to explore and grow. I was able to do exactly what I wanted to do in graduate school. You believed in me even when I failed to, you taught me what it means to be a scientist with integrity. I'm truly grateful for the journey.

To Jeff, thank you for teaching me how to code and how to learn math. I'll always code with a "Varner style" and I'm proud of it. Thank you for putting up with me busting into your office for questions constantly, I truly enjoyed learning from you.

To fellow Daniel lab members, thank you for being my office mates and my friends, and always spares me a piece of pizza after lab meetings. Angela and Hunglun, thank you for being my best friends and my emotional supports at Cornell. I couldn't have accomplished half of my research without you two listening and helping me trouble shooting it along the way, although my work isn't even related to yours! Rohit, thank you for being my friend, tennis coach and running buddy, I miss your stoic face and dry humor dearly.

To David, thank you for being the little brother I've been missing in life. Thank you for taking turns with me to being the crazy twin. My Ithaca memories were full of fun adventures and laughters because of you. My favorite moments of graduate school will always be us drinking white wine on the slope at sunset. I miss you and Ithaca, but no matter how far apart we are, we are always family.

To Yan, Lulu and Katerina, thank you for being my students! I'm extremely proud of you all! You taught me how to teach and guide the growth of young scientists, and you offered me tremendous help. I couldn't have published my last two papers without your hard work. You all have grown from my students to my colleagues, whom I respect and heavily rely on. I hope you continue your journeys in science with the your immense

passion and dedication.

To Rob, Thank you for showing up in my life and being there to catch me when my world collapsed. I couldn't have made it to the end of my Ph.D. without your love and support. You brought laughter and joy into my life. I have so much more to look forward to in my future because of you.

To Jenny, my gratitude towards you is beyond what words can convey. You were the most beautiful thing that has ever happened to my life. You pulled me out of the swamp of self-destruction over a decade ago. Then you traveled across the world to become my only family in my lonely quest for science. For 11 years, you played the roles of my dog, my child, my mom, my sister and my best friend. You were there through all my self-doubts, all the frustration, anger, loneliness, anxiety and fear. For your whole life, you never failed to comfort me. Thank you for always quietly sitting next to me and putting your big paw on my shoulder when I was upset. Thank you for offering the warmest, sturdiest hugs. Thank you for having the biggest, the most infectious smiles. Thank you for your passion for life and adventures. Thank you for being a spunky sassy lady. Thank you for fighting cancer bravely for 5 years, I know you were worried about me and didn't want to leave me alone. I lost you several months after your 11th birthday, but I know that I'll never be alone again. You're forever a part of me —The part that loves nature, the part that's patient to students, the part that hates wasting food, the part that cherishes every living moment, the part that values families and friends, and the part that is forever a mom —forever your mom.

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . v
Acknowledgements . . . vi
Table of Contents . . . viii
List of Tables . . . xi
List of Figures . . . xii

1 Introduction **1**
1.1 The brief history of synthetic biology . . . 1
1.2 RNA synthetic biology . . . 3
1.3 Mathematical framework of biological networks . . . 5
1.4 The advantages of cell-free systems . . . 6
1.5 Building theoretical foundation for RNA synthetic biology . . . 8
1.5.1 Chapter 2: Building the first predictive model for RNA circuitry 8
1.5.2 Chapter 3: Apply the predictive model to design and engineer complex RNA circuitry . . . 9
1.5.3 Chapter 4: Define the design principle for transfer function manipulation with sRNA . . . 10

2 Generating effective models and parameters for RNA genetic circuits[1] **12**
2.1 Abstract . . . 12
2.2 Introduction . . . 13
2.3 Results and Discussion . . . 18
2.3.1 Model Derivation . . . 18
2.3.2 Parameter Estimation Through Sensitivity Analysis-Based Experimental Design . . . 24
2.3.3 Model Prediction . . . 31
2.3.4 Understanding TX-TL Batch Variation Using Model Parameterization . . . 33
2.4 Conclusion . . . 37
2.5 Methods . . . 40
2.5.1 Sensitivity Analysis to Design Parameterization Experiments . . 40
2.5.2 Parameter Estimation . . . 42
2.5.3 Plasmid construction and purification . . . 44
2.5.4 TX-TL extract and buffer preparation . . . 44
2.5.5 TX-TL Experiment . . . 45

[1] This work was originally published in ACS Synthetic Biology and has been reproduced here with permission. Hu, C. Y., Varner, J. D., & Lucks, J. B. (2015). Generating Effective Models and Parameters for RNA Genetic Circuits. ACS Synthetic Biology, 4(8), 914926. http://doi.org/10.1021/acssynbio.5b00077

3 Engineering a Functional small RNA Negative Autoregulation Network with Model-guided Design [2] **47**
3.1 Abstract . . . 47
3.2 Introduction . . . 47
3.3 Results and Discussion . . . 51
3.3.1 Using simplified computational models to guide NAR network design . . . 51
3.3.2 Optimization of RNA parts for NAR network construction . . . 54
3.3.3 A quantitative model accurately predicts NAR network function in TX-TL . . . 58
3.3.4 The sRNA transcriptional NAR network functions in *E. coli* . . . 67
3.3.5 Noise reduction properties of sRNA transcriptional NAR networks 69
3.4 Conclusion . . . 70
3.5 Method . . . 72
3.5.1 Plasmid construction and purification . . . 72
3.5.2 TX-TL experiments . . . 73
3.5.3 Network/Model Simulations . . . 74
3.5.4 Parameterization experiment design . . . 74
3.5.5 Parameter estimation . . . 75
3.5.6 Strains, Growth Media and *in vivo* Gene Expression Experiments 76
3.5.7 Response-time Calculation . . . 77
3.5.8 Signal Noise Analysis . . . 79
3.6 Acknowledgement . . . 79

4 Uncover the design principle of generic transfer function manipulation utilizing small RNA regulators [3] **80**
4.1 Abstract . . . 80
4.2 Results and Discussion . . . 84
4.2.1 Understand the transfer functions of STAR hybrid activators . . . 84
4.2.2 STARs equips transciprtional inducible systems with a tuning knob . . . 89
4.2.3 STAR amplifies dynamic range . . . 94
4.2.4 STAR sequesters also impact transfer function and dynamic range 100
4.2.5 Broad applications and limitations . . . 102
4.3 Conclusion . . . 106
4.4 Material and Methods . . . 108
4.4.1 Model simulation . . . 108
4.4.2 TX-TL experiments . . . 108
4.4.3 Parameter estimation . . . 109

[2] This work was originally published in ACS Synthetic Biology and has been reproduced here with permission. Hu, C. Y., Takahashi, M. K., Zhang, Y., & Lucks, J. B. (2018). Engineering a Functional Small RNA Negative Autoregulation Network with Model-Guided Design. ACS Synthetic Biology, 7(6), 15071518. http://doi.org/10.1021/acssynbio.7b00440

[3] This work is being prepared as a manuscript for review.

4.4.4 Strains, growth media and *in vivo* gene expression experiments . 110

5 Conclusion and Perspectives 112
5.1 Conclusion 112
5.2 Future perspectives 113

A Appendix of Chapter 2: Generating effective models and parameters for RNA genetic circuits 115

B Appendix of Chapter 3:Engineering a Functional small RNA Negative Autoregulation Network with Model-guided Design 131
B.1 Supplementary Note: ODEs that model individual parameterization experiments 131
B.2 Methods for sensitivity analysis and parameter identifiability 133

C Appendix of Chapter 4: Uncover the design principle of generic transfer function manipulation through small transcription activation RNAs (STARs) 152
C.1 Parametrization on Three Pairs of STARs 152
C.1.1 Model Species and Parameters 152
C.1.2 Equations 154
C.1.3 Parametrization Methods 154
C.1.4 Sensitivity Analysis 157
C.1.5 Simulation Results 158
C.1.6 STAR Pairs and Parametrized Values 159
C.2 *In vivo* Simulation with Three Pairs of STARs 161
C.2.1 Model Species and Parameters 161
C.2.2 *In vivo* Equations 161
C.2.3 Parameter Values used for *in vivo* Simulation 163
C.3 Understanding Leak coefficient l_x of transcription factor mediated activator 164
C.3.1 Positive Activation 164
C.3.2 Repressor sequestrating activation 166
C.4 Degradation Model for sRNA 169
C.4.1 Model Species and Parameters 169
C.4.2 Equation 170
C.5 STAR Sequester impacts transfer function and dynamic range 173
C.5.1 Model Species and Parameters 173
C.5.2 Equations 174

Bibliography 176

LIST OF TABLES

2.1 Model Parameters 22
2.2 Model Species 23

3.1 NAR model parameters 59
3.2 NAR model species 59

4.1 Species involved in hybrid inducible system 86
4.2 Species involved in hybrid inducible system 86

A.1 Important DNA sequences 124
A.2 Plasmids used in this study. Sequences in the plasmid architecture can be found in Table A.1 128
A.3 Estimation constraints 129
A.4 Estimated parameter mean values 130

B.4 Important DNA sequences 136
B.5 Plasmids used in this study. (Sequences in the plasmid architecture can be found in Table B.4) 140
B.1 Parameters guesses from previous work[48] 145
B.2 All parameters involved in parameterization procedure 150
B.3 Model Species 151

C.1 Species involved in STAR/target Circuit 152
C.2 Parameters involved in STAR/target Circuit 153
C.3 Estimated Parameter Values 159
C.4 Species involved in hybrid inducible system 161
C.5 Parameters involved in hybrid inducible system 162
C.6 Parameters values used in hybrid inducible system 163
C.7 Species involved in STAR-hybrid system 169
C.8 Parameters involved in STAR-hybrid system 170
C.9 Species involved in STAR-hybrid system 173
C.10 Parameters involved in STAR-hybrid system 174

LIST OF FIGURES

1.1 Mechanism of antisense sRNA mediated transcriptional regulation. (A) A schematic of the sRNA repressor mechanism. In the absence of antisense sRNA repressor, the anti-terminator sequence prevents the terminator formation and allows downstream transcription (ON). When present, the antisense sequesters the anti-terminator sequence and allows terminator formation, preventing downstream transcription (OFF). (B) Schematic of the STAR mechanism. In the absence of a STAR antisense, an intrinsic terminator is formed in the sense target RNA, preventing transcription elongation (OFF). In the presence of the STAR antisense, the 5' intrinsic terminator stem is sequestered by the STAR antisense, allowing downstream transcription by RNA polymerase (RNAP). Figure from Chappell et al.[20] 4

1.2 A schematic of a transcription event described by one simple Hill function. A repressor, X, is a transcription factor protein that decreases the rate of mRNA transcription when it binds to the promoter. The signal, S_x, increases the probability that X is in its active form X^*. X^* binds a specific site in the promoter of gene Y to decrease transcription and production of Y protein. One ODE is adequate to describe the entire event. Adapted from *Alon, Uri, (2006) —An Introduction to systems biology*[1]. . . . 6

1.3 Schematic describing the process of arraying synthetic gene networks on paper using printed arrays. Figure from *Pardee et al* [84]. 7

2.1 Schematic of the model development and parameterization process. (A) A 3-level, DNA-encoded RNA transcriptional cascade (colored circles) is composed of two orthogonal sRNA repressor/attenuator pairs. The RNAs are configured in a double repression cascade, with the final output being the transcription of a target gene that encodes a translated protein. The complex mechanism of RNA transcriptional repression can be described by coarse-grained ordinary differential equations (ODEs) with a handful of unknown parameters (red). (B) Parameterization experiments can be designed based on a parameter sensitivity analysis of the model equations. This analysis identifies which parameters can be estimated from a particular experimental design. These experiments can then be performed to estimate the indicated parameters. This process can then be iterated until all parameters are estimated, resulting in distributions of parameters that accurately model the desired genetic circuitry. 17

2.2 Governing Equations for a 3-level sRNA transcriptional cascade (A) Schematic of the cascade. Level 1 (L1) plasmid expresses attenuator 1 (A_1) controlling SFGFP transcription; Level 2 (L2) expresses attenuator 2 (A_2) controlling the transcription of tandem copies of sRNA repressor 1 (R_1); Level 3 (L3) expresses sRNA repressor 2 (R_1). Concentrations of DNA templates used in TX-TL reactions for part B are listed beside the levels. (B) Representative fluorescence signal time trajectories in TX-TL reactions containing three different combinations of DNA templates from the transcriptional cascade in part A. L1 alone leads to a high rate of SFGFP production (blue line); L1+L2 results in results in a reduced SFGFP production rate (red line) due to R_1 repressing A1; L1+L2+L3 (purple line) results in a higher SFGFP production rate than just L1+L2 due to the double negative inversion of the full cascade. (C) Schematic (left) of the mechanistic steps of the cascade in part A that are captured by the governing equations (right). The equations model the tandem copies of R_1 on L2 as one repressor. 20

2.3 Sensitivity matrices for parameter identification experiments. Four parameterization experiments were designed based on sensitivity analysis to estimate all thirteen unknown parameters in our model (Figure 2.2). For each experiment, the constructs used are shown above the SFGFP portion of the calculated sensitivity matrix for that experiment. Parameters are numbered according to Table 2.1. Red/blue indicates high/low sensitivity, respectively. Note that time varying changes in parameter sensitivity indicates portions of the trajectories that are influenced by each parameter. Experiments were designed in order from (A) to (D), with parameters identified by previous experiments marked as grey rows. Since the same promoter was used on all constructs, β_1 and β_2, determined by multiplying β_m by an appropriate factor based on DNA template concentrations, are absent from the sensitivity analysis. (A) 0.5nM of L1 plasmid alone is able to identify five parameters: β_1, k_E, d_m, α and k_I. (B) 0.5nM of L1 and 4nM of L2 is able to identify four parameters: K_1, d_1, r_{m1} and P_t. (C) 0.5nM of L2 (Table A.2) and 4nM of L3 is able to identify three parameters: K_2, d_2, r_{m2} (D) 0.5nM of L1, 4nM of L2, and 14nM of L3 is able to identify the last parameter, p_c. The sensitivity color scale was changed in D to aid in visualization. . 27

2.4 Validation of model simulations of parameter estimation experiments. Comparison of experimental trajectories of SFGFP fluorescence in TX-TL experiments (black dash lines) with simulated model predictions. Model simulated trajectories were generated by performing 1,000 simulations with parameters drawn from the set of 10,000 determined from the estimation procedure (see Methods). Experimental and model trajectories were normalized by the maximum observed experimental fluorescence (see Methods). The mean simulated trajectory (red line) is shown within 95% confidence intervals (blue region). The schematic of each experiment is shown in the upper left corner of each plot corresponding to the experiments in Figure 2.3. 30

2.5 Comparison of experimental trajectories of SFGFP fluorescence in TX-TL experiments (black dash lines) with simulated model predictions. Model simulated trajectories were generated by performing 1,000 simulations with parameters drawn from the set of 10,000 determined from the estimation procedure (see Methods). Experimental and model trajectories were normalized by the maximum observed experimental fluorescence (see Methods). The mean simulated trajectory (red line) is shown within 95% confidence intervals (blue region). The schematic of each experiment is shown in the upper left corner of each plot. (A) Two level concentration prediction that varies the L2 plasmid concentration from Figure 2.3B. (B) Three level concentration prediction that varies the L3 plasmid concentration from Figure 2.3D. (C) Two level tandem attenuator prediction. The experiment contains 0.5nM of a modified L1 plasmid expressing 2 tandem copies of A_1 in front of SFGFP (L1T, Table A.2), and 4nM of the L2 plasmid. (D) Three level tandem attenuator prediction containing the same constructs as in (C) with an additional 14nM of L3 plasmid. (E) Two level swap prediction. The experiment contains 0.5 nM of a modified L1 plasmid that expresses A_2 in front of SFGFP (PL1, Table A.2), and 8nM of a new L2 plasmid expressing A_1 followed by R_2 (PL2, Table A.2). (E) Three level swap prediction containing the same constructs as in (D) with an additional 14nM of a new L3 plasmid that expresses R_1 (PL3, Table A.2). 34

2.6 Parameter distributions from two independent TX-TL batches. Estimated parameters from batch A (blue) (Figure 2.4) and batch B (red) (Figure A6). Histograms are composed of 10,000 sets of parameters fit from the parameterization experiments performed in each batch. 35

3.1 A simple model of the proposed sRNA transcriptional NAR construct uncovers potential challenges and provides design guidelines for network construction. (A) The transcriptional NAR network motif and proposed sRNA implementation. The pT181 attenuator regulates the transcription of N copies of its own repressor RNA (repressor, R) and a reporter gene (G), each insulated from each other by self-cleaving ribozymes (triangles). Once transcription is activated, initially RNA polymerase is allowed to transcribe the repressor RNA-reporter gene construct. As concentration of repressor RNA increases, it binds to the attenuator, which leads to intrinsic terminator formation and repression of transcription of the repressor RNA-reporter gene construct. (B) A simplified two-component model that describes the synthesis and degradation of the repressor (R) and reporter gene (G) provides insight for network construction. The dynamic trajectories were simulated by this model with parameters adapted from a previous study (Table B.1)[48]. The simulation suggests that a single copy of repressor RNA ($N = 1$) would be insufficient to significantly reduce network response time, while tandem copies of repressor RNA ($N = 2$) would. 53

3.2 Optimization of RNA parts for network construction. (A) Average steady-state fluorescence (FL/OD) of *E. coli* TG1 cells measuring repression of a construct consisting of the pT181 attenuator controlling SFGFP expression. The pT181-SFGFP construct was transformed with a control plasmid (-), or constructs that expressed variations of the repressor RNA. Variations in the terminator included downstream of the repressor tested the impact of the size of the terminator (T1 vs. T2), with L3S3P2134 (T2) being smaller than TrrnB (T1). Variations in the size of the repressor RNA repressor (two hairpins vs. one hairpin) tested the impact of including reduced regions of the repressor RNA on repressive function17. Other variations tested the impact of using the sTRSV ribozyme (triangles) for separating RNA parts and including multiple copies of the repressor. Error bars represent standard deviations over nine biological replicates. (B) Orthogonality matrices illustrating that the single hairpin repressor RNA followed by terminator L3S3P21 achieves adequate repression but compromises orthogonality (left). Matrices were generated by measuring the fluorescence of cells transformed with different combinations of the wild type (WT) or specificity mutated (MUT) pT181 attenuator-SFGFP/repressor expression constructs in the repressor configurations indicated by the schematic. Adding a ribozyme in between the single hairpin repressor RNA and terminator L3S3P21 restores the orthogonality (right). Each box in the matrices represents % repression of cells with a repressor RNA expression plasmid compared to cells with the control plasmid. Repression is presented by a color scale in which 100% is blue and 0% is white. Repression % was calculated from the average over nine biological replicates. 57

3.3 Schematic (left) of the mechanistic steps of the sRNA transcriptional NAR network captured by the governing equations (right). Molecular implementations that highlight key rate constants and interactions are shown below each network schematic. N refers to the number of tandem repressors included in each construct. The equations model the tandem copies ($N = 2$) of repressor R_1 as a single new repressor, R_2. This effective model also used different degradation rate constants, d_{mN}, for each repressor species. Parameter and Species descriptions can be found in Table 1 and 2, respectively. The only difference between the negative autoregulation (NAR) and control (CTRL) constructs is the use of a mutated attenuator target region that disrupts repression, modeled as a change in repression coefficient K_{NC}, describing a weak non-cognate repression caused by crosstalk. 60

3.4 Sensitivity matrices for parameter identification experiments. Five parametrization experiments were designed on the basis of sensitivity analysis to estimate all 16 unknown parameters in our model (Supplementary Note Equations B.1.1-B.5.2). For each experiment, schematics of the constructs used are shown above the SFGFP portion of the calculated sensitivity matrix for that experiment, with DNA concentrations indicated. Parameters are numbered according to Table 3.2. Red/blue indicates high/low sensitivity, respectively. Note that time varying changes in parameter sensitivity indicates time windows of the trajectories that are influenced by each parameter. Experiments were designed in order from (A) to (E), with parameters identified by previous experiments marked as gray rows. (A) 0.5 nM of reporter plasmid alone is able to identify four parameters: β_m, k_t, α and d_m. (B) 0.5 nM of reporter and 4 nM of repressor plasmid is able to identify five parameters: K_1, d_1, r_{m1}, β_R and P_t. (C) 0.5 nM of reporter and 4 nM of plasmids containing repressor with two tandem repressors (separated by a ribozyme) is able to identify three parameters: K_2, r_{m2} and d_2. (D) 0.5 nM of reporter with a mutated PT181 attenuator and 4 nM of repressor plasmid is able to identify one parameter K_{C1}. (E) 12 nM of a plasmid with MG as reporter identifies the last three parameters: r_b, β and d_{MG} 64

3.5 Model prediction and experimental characterization of the sRNA transcriptional NAR constructs in TX-TL. (A) Schematics of the NAR and control constructs designed for TX-TL using malachite green (MG) as a network reporter. (B) and (C) Comparison of experimental trajectories of MG fluorescence in TX-TL experiments (solid circles with error bars) with simulated model predictions (shaded regions) for single repressor NAR ($N = 1$) and double repressor NAR ($N = 2$) networks, respectively. Experimental error bars represent standard deviations of nine individual TX-TL reactions. Model simulated trajectories were generated by performing 1000 simulations with parameters randomly drawn from the set of 10,000 determined from the estimation procedure (see Methods). Predicted and measured response times for all four trajectories are shown in the upper left corner. 66

3.6 Demonstration and characterization of a functional sRNA transcriptional NAR network *in vivo*. (A) Schematics of the NAR and control constructs designed for *in vivo* testing using yem-GFP as a network reporter. (B) Time course of average fluorescence (FL/OD) collected from *E. coli* TG1 cells containing the double repressor NAR (blue) or control (red) construct over a five-hour period. Shaded regions represent standard deviations collected at each time point for nine biological replicates. Calculated response times (see Methods) indicate a functional NAR network that decreases the response time with respect to the control. (C) Flow cytometry histograms of a single time point measurement after 5 hours, tested with the double repressor NAR (blue) and control constructs (red) *in vivo*. Individual distributions are overlaid on each other with different shades of blue and red. The ratio of relative variance (RV) and coefficient of variation (CV) was calculated for each distribution and averaged. The averages of the ratio of (RV_{NAR}/RV_{CTRL}) and (CV_{NAR}/CV_{CTRL}) are shown in the upper right corner with errors representing the standard deviation of these calculated ratios. 68

4.1 Understanding the transfer function shift caused by STARs. (A) The schematics of the STAR hybrid operon. The inducible promoter regulates the expression of STARs; STARs then regulates the expression of GFP. (B) Re-plot of an observation from previous work[22]. Three variants of STAR hybrid inducible operon presented 3 different transfer functions. (C) Transfer functions simulated based on estimated key STAR parameters. They largely resemble the observed transfer function from B. (D)Key STAR parameters estimated based with STAR circuit dynamics in TX-TL, see Appendix C1 85
4.2 STAR hybrid operon provides an tuning knob to manipulate the otherwise static transfer function. (A) The transfer function of a *Xylose* operon on high copy ColE1 backbone, with a two fold dynamic range. (B) The entire transfer function moved to a medium copy backbone P15A. The operon became non-inducible. (C) The STAR hybrid operon leaves a constitutive promoter open for substitution (orange). Three weaker Anderson promoters were titrated in the configuration, the transfer function gradually moves down, reduces signal leak of the operon.(D) The dynamic ranges of these four transfer functions stay relatively constant 89
4.3 STARs amplify dynamic range of inducible systems (A) Dynamic range amplification was observed when configured with high strength constitutive promoter (orange promoter in 4.2C). Dynamic range was amplified from 2 fold to 6 folds when reporter gene was expressed via a medium copy plasmid P15A (red),further amplified to 10 folds when expressed via a high copy plasmid ColE1 (maroon). (B) Simulation results from Appendix C note 4. Simulated dynamic range decreased When the hybrid operon was modeled with constant degradation rate (hollow circles). Simulated dynamic range showed improvement with the same set of parameters when modeled with constant RNase pool. (C) the schematics of STAR mediated dynamic range amplification (D) Parameters of systems impact dynamic range amplification effects. (E) Dynamic range amplification induced by STAR 4971. (F) Dynamic range amplification induced by STAR 5809. 95
4.4 STAR sequesters impact transfer functions (A) Schematics of the hybird operon configured with STAR sequester. (B) Simulated transfer function shift when hybrid operon configured with STAR sequester. The on state have a larger signal deduction than the off state due to RNA degradation slowdown (C) Experimentally collected transfer functions. Comparing xylose operon (grey, 2 fold),xylose STAR hybrid (red solid dots 10 fold) and hybrid operon configured with sequester (red hollow circles 14 fold). The experimental result agreed with simulation that the ON state signal suffered a larger deduction than the OFF state. 100
4.5 Broad application and limitation of STAR as dynamic range amplifier. (A) STAR improves a house-assembled Lac operon from 4 folds to 12 folds. (B) STAR improves a theophylline inducible transcriptional riboswitch from 2.7 fold to 5 folds. (C) The Xyl-STAR hybrid operon's dynamic range amplification effect in TG1 was not observed in MG1655. (D) STAR did not improve the dynamic range of xylose operon in the Keio parent strain. Knocking out either MazF or RelA recover the amplification effect. 102

A.1 Model components of the 3-level cascade. (A) Transcriptional repression governed by sRNA repressors. Presence of R_2 causes attenuator A_1 upstream of R_1 to fold into a terminator structure thus halting transcription of R1. This complex event can be described by a first order Hill function[64], characterized by a repression coefficient, K_2,which has units of concentration. K_2 defines the concentration of active R_2 needed to achieve significant repression of R_1[1]. (B) Crosstalk between repressor R_2 and attenuator A_1 can be described as an OR gate module with two signaling species[71]. Here p_c is the percentage crosstalk due to R_2 compared to the repression caused from the cognate sRNA repressor R_1. (C) Tandem attenuators can be modeled by multiplying Hill functions together[64].(D) The effect of autotermination from a single attenuator is modeled by multiplying the transcription rate by a factor of $(1 - P_t)$, where Pt is the probability of auto-termination 119

A.2 Qualitative validation of model governing equations. (A) Cross-talk qualitative validation, generated by equations A.5-A.8. As cross-talk becomes stronger, the SFGFP signal is more repressed. (B) Tandem attenuator and maturation delay validation, generated by equations A.23-A.30. Both trajectories resulted from L1+L2, one with a single attenuator on L1, and one with two tandem attenuators on L1. When two attenuators are in tandem, the repression effect caused by the same repressor RNA concentration is stronger. These curves also show the characteristic bending of the trajectories at later times due to the maturation delay we factor into the model, which matches experimental characterization (Figure A.4). (C) Auto-termination validation, generated by equations A.13-A.16. Compared with no auto-termination (left two bars), 20% auto termination (right two bars) causes end point signals at 100mins to drop for both levels (L1 and L1+L2) as expected. 120

A.3 Model inspired findings on pre-incubation of TX-TL reactions. (A) Best fit of the model generated L1 trajectory with equations A17-A22 (red line) and experimentally collected SFGFP trajectories (blue dash) from TX-TL reactions containing 0.5nM L1 with no pre-incubation. (B) GFP production rates derived from part A. Red line corresponds to the red model trajectory and blue dashes correspond to blue experimental trajectories. Parts A and B together show that there is a 20 minute delay in SFGPF fluorescence signal after the start of the reaction, which our model is not capable of capturing. (C) Best fit of the model generated L1 trajectory with equations A17-A22 (red line) and experimentally collected SFGFP trajectories (blue dash) from TX-TL reactions containing 0.5nM L1 with a 20 minute pre-incubation of the extract and buffer at 37°C before adding DNA at time 0. (D) GFP production rates derived from part C. Red line corresponds to the red model trajectory and blue dashes correspond to blue experimental trajectories. Parts C and D together show that a 20 minute pre-incubating of the TX-TL extract/buffer at 37°C eliminates the delay at the beginning of the reaction. Using experimental pre-incubation, our model is able to qualitatively capture the observed SFGFP trajectories. 121

A.4 Model inspired findings on sRNA maturation delay (A) Best fit of the model generated L1+L2 trajectory with equations A17-A22 (red line) and experimentally collected L1+L2 trajectories (blue dashes) with a 20 minute pre-incubation of the extract and buffer at 37°C before adding DNA at time 0. (B) GFP production rates derived from part A. Red line corresponds to the red model trajectory and blue dashes correspond to blue experimental trajectories. Parts A and part B together show that there is a drop in SFGFP production rate after the first 20 minutes in the single repression construct, which equations A17-A22 cannot capture. 122

A.5 Estimation of K_{1s} for cascade swap prediction. (A) The two-level parameterization experiment (black dash lines) was performed to estimate a single parameter K1s to predict trajectories for a swapped cascade circuit (Figure 2.5E,F). The mean simulated trajectory (red line) is shown within 95% confidence intervals (blue region) using the newly estimated K1s values (B) Comparison of the estimated repression coefficients between single antisense copies of R_1 (Red) and R_2 (Blue) (K_{1s} and K_{2s}, respectively). K_{2s} is lower than K_{1s} indicating that R_2 shows slightly stronger repression as has been observed previously[110, 108]. 122

A.6 Parameter estimation validation on TX-TL batch B. Comparison of experimental trajectories of SFGFP fluorescence in TX-TL experiments using batch B of extract (black dash lines) with simulated model predictions. Model simulated trajectories were generated by performing 1,000 simulations with parameters drawn from the set of 10,000 determined from the estimation procedure (see Methods). Experimental and model trajectories were normalized by the maximum observed experimental fluorescence (see Methods). The mean simulated trajectory (red line) is shown within 95% confidence intervals (blue region). The schematic of each experiment is shown in the upper left corner of each plot corresponding to the experiments in Figure 2.3. . 123

A.7 Batch-to-batch difference Parameterization experimental trajectories comparison between batch A (blue) and batch B (red). Shaded regions represent the average plus/minus the standard deviation of experimental trajectories from Figure 2.4 (Batch A) and Figure A.5 (Batch B). 124

B.1 (A) Natural repressor-RNA transcriptional repression in the pT181 attenuator. Transcription of a downstream gene is undisrupted when the repressor RNA is absent. When the repressor RNA is present, it interacts with the attenuator RNA to form a transcriptional terminator hairpin. Therefore, the transcription of the downstream gene is turned off. (B) Schematics of basic sRNA repression constructs testing both cognate and non-cognate (orthogonal) regulation. (C) In vivo test of 5 new terminators compared against TrrnB in the context of the sRNA repressor. Error bars represent standard deviations over nine biological replicates. L3S3P21 was chosen for the rest of this work. (D) *Invivo* orthogonality test of double (tandem) repressors. Adding a ribozyme (triangle) between the terminator and repressor construct restores orthogonality. Error bars represent standard deviations over three biological replicates. 135

B.2 Re-plots of previous findings on pT181 sRNA repression strength and degradation rates. (A) Re-plot of in vivo attenuator repression characterization from *Lucks et al*[64]. The repression efficiency of pT181 repressor antisense (R) is reduced when an attenuator sequence is placed upstream (att-R). Using two copies of repressors in tandem (att-2R) increases the repression strength slightly. (B) Re-plot of degradation parameter distributions for different repressor constructs determined from cell-free TX-TL experiments in *Hu et al*[48]. The degradation rate of the bare pT181 repressor sRNA (R, red) is faster than the degradation rate of the att-2R version (blue). 136

B.3 Characterization of malachite green and SFGFP expression constructs. (A) End point fluorescent measurement at the end of a 2-hour TX-TL reaction with a constitutive promoter driving malachite-green aptamer expression (MG) compared with a version containing a repressor-ribozyme sequence in between the promoter and the MG aptamer (R-MG). This result indicates some reduction in MG transcription caused by the presence of the repressor-ribozyme sequence. The ribozyme is indicated by a triangle in the schematic. Error bars represent standard deviations of 3 technical replicates. (B) Two-hour TX-TL time course trajectories of constructs containing the wild type pT181 attenuator followed by an SFGFP coding sequence (WT-att), compared with a mutated version of the pT181 attenuator (Mut-att) followed by an SFGFP coding sequence. The effective model simulates these two constructs with identical ODEs. Therefore, the two trajectories are expected to overlay each other if the DNA qualities of these two constructs are comparable. The result confirmed that this modeling approach is valid. Error bars represent standard deviations of 9 technical replicates. . 144

B.4 Validation of model simulations of parameter estimation experiments. Comparison of experimental trajectories of SFGFP (or MG) fluorescence in TX-TL experiments (black dashed lines) with simulated model predictions. Model simulated trajectories were generated by performing 1000 simulations with parameters drawn from the set of 10,000 determined from the estimation procedure (see Methods). Experimental and model trajectories were normalized by the maximum observed experimental fluorescence of the first experiment in (A). The mean simulated trajectory (red line) is shown within 95% confidence intervals derived from the range of simulated trajectories (blue region). The schematic of each experiment is shown in the upper left corner of each plot corresponding to the experiments in Figure 3.4 and equations in Supplementary Note B.1. 145

B.5 To estimate the response time of dynamic trajectories generated by constructs in TX-TL reactions, we used a least squares method to approximate when each trajectory reaches steady state (see Methods). Since the raw data contains experimental noise within individual trajectories, steady state approximation is a difficult task. To overcome this, we fitted the data with our model and used the best fit trajectory to estimate the response time. (A) We searched for the best trajectory for each individual experimental replicate based on best R2 values. Sample fitting trajectories and experimental data from each experiment are shown. (B) All fitted trajectories and experimental data from the single repressor NAR vs. control constructs. (C) All fitted trajectories and experimental data from the double repressor NAR vs. control constructs. Response time of each experimental replicate (shown in Figure 3.5) was calculated independently following the least squared method (See Methods). 146

B.6 Parts tested for the sRNA transcriptional NAR networks designed for in vivo experiments. (A) Testing yem-GFP in the context of the sRNA repressor. Expression was characterized from constructs containing a constitutive promoter followed by a pT181 attenuator and either GFP or yem-GFP with a control plasmid (-), or a plasmid encoding a cognate sRNA repressor (R,+). Results showed an improvement in repression strength when yem-GFP was used. Error bars represent standard deviations over nine biological replicates. (B) Using AHL to titrate expression of an attenuator construct. The attenuator-yem-GFP construct was placed behind the AHL inducible promoter pLux. Construct expression was measured after induction with a range of AHL concentrations after 5 hours, and compared to a constitutively expressed construct (J23119). The results confirmed a range of AHL induction levels of this repressor expression construct. Later in vivo experiments were performed with 100nM of AHL based on this result. Error bars represent standard deviations over nine biological replicates. (C) Repression efficiency of the single repressor construct used in the single repressor NAR networks. The repression strength of a single repressor flanked by ribozymes (triangle) versus the same construct with a mutant attenuator placed upstream. The present of the mutant attenuator significantly reduces the repression strength. Testing the repressor construct against a mutant attenuator target region (red bars) showed that the repression inefficiency was not due to crosstalk, but is rather due to the attenuator causing reduced sRNA transcription (see Figure B.2A). Error bars represent standard deviations over three technical replicates. 147

B.7 Characterization of double repressor sRNA transcriptional NAR networks in vivo with normalized data. (A) Schematics of the double repressor NAR and control constructs designed for in vivo testing using yem-GFP as a network reporter. (B) Normalized trajectories collected from *E. coli* TG1 cells containing the double repressor NAR (blue) or control (red) construct over a five-hour period. Each experimental replicate from Figure 6B was normalized to its steady state value to show that NAR speeds up network response time. 148

B.8 Characterization of single repressor sRNA transcriptional NAR *in vivo* . (A) Schematics of the single repressor NAR and control constructs designed for *in vivo* testing using yem-GFP as a network reporter. (B) Four replicate fluorescence (FL/OD) trajectories collected from *E. coli* TG1 cells containing the single repressor NAR (blue) or control (red) construct over a five-hour period. In this case, no clear speed up in response was observed. 149

C.1 Six parametrization experiments were designed on the basis of sensitivity analysis to estimate all 22 unknown parameter values in our proposed model. Red and blue indicated high and low sensitivity, respectively. The first construct level (A) contains only a set concentration of the 4970 target DNA, representing the off level of the 4970/4971 STAR/target pair. The second construct level (B) contains both the 4971 STAR DNA and the 4970 target DNA at set concentrations, representing the on level. Likewise, construct levels 3 (C) and 4 (D) contain the 5808 target DNA and the 5808/5809 STAR/target DNAs, respectively. Construct levels 5 (E) and 6 (F) contain the 5816 target DNA only and the 5816/5817 STAR/target DNAs, respectively. . . . 157

C.2 Comparison of experimental trajectories of SFGFP fluorescence in TX-TL experiments (black dashed lines) with simulated model predictions. Model simulated trajectories were generated by performing 100 simulations with parameters drawn from the set of 10,000 determined from the parameter estimation procedure. Experimental and model trajectories were normalized by the maximum experimental fluorescence of the first experiment (A). The mean simulated trajectory (red line) is shown within 95% confidence intervals (shaded blue region). Plots (A) and (B) 2 correspond to 4970 and 4971 pair. Plots (C) and (D) correspond to 5808 and 5809 pair. Plots (E) and (F) correspond to 5816 and 5817 pair. 158

CHAPTER 1

INTRODUCTION

1.1 The brief history of synthetic biology

Synthetic biology is an engineering discipline. Although still in its infancy, it holds tremendous promise for applications in many aspects of technology [25]. The central goal of synthetic biology is to engineer non-existing biological systems and reprogram living cells to perform tasks or create chemicals. The potential ability of harness the power of biology would entirely transfer the way we interact with medicine, environment and energy[3].

The root of synthetic biology can be traced back over half a century ago when Fancois Jacob and Jacques Monod in 1961 published the first study on the *Lac* operon in *E.coli* [73]. This study marked the discovery of synthetic biology's first regulatory component, which was extensively used decades later to form molecular regulatory systems. In the decade of 1970-1980, genetic manipulation became widespread in microbiology research because of the development of molecular cloning techniques. However in this stage, genetic engineering was missing the necessary knowledge to create diverse and in-depth biological systems we found in natural microorganisms. This missing piece of the puzzle was eventually discovered in the 1990s, when automated DNA sequencing and improved computational tools enabled complete microbial genome sequencing. This major development created the field of systems biology, which studies complex cellular networks and discovers the functional modules of these networks [51]. The two work that marked the birth of synthetic biology was the successful engineering of genetic toggle switch [37] and repressilator [31]. They also envisioned that the dynamics of simple genetic regulatory circuits can be described using simple mathematical mod-

els. This inspired biologists to draw analogy from electrical circuit engineering when creating synthetic biological networks, which are often refereed to as ”circuits” today.

In the following years nearly two decades,the growth of synthetic biology accelerated by the breakthrough in next-generation sequencing and advance computing power. During this time, numerous synthetic biological ”circuit components” were created, modularized and integrated into cellular networks. First of all, Hsien's discovery of green fluorescent protein and his sequential work that expanded the library of fluorescent proteins [96] revolutionized biological assays. These fluorescent proteins play the role of ”light bulbs” in synthetic circuits, provides a fast and reliable method for circuit measurement and analysis. At the same time, regulatory systems as the heart of synthetic circuits has also been one of the the main focuses of the field. From the traditional small molecule sensing regulators like *Lac* operon and *Tet* operon, scientists have developed systems that senses light [62], temperature [94], heavy metal toxins [115] and more. Many aspects of genetic expression has also been engineered. Genetic parts are engineered to regulate transcription, translation and post-translation [87]. These parts have been wired into complete sets of logic gates [112], multiple input logic cascades [74], event-counting circuits[34], edge-detector circuits[107], just to name a few.

Synthetic biology today has grown to a new stage where scientists are moving from engineering stand-alone circuits to complex and sophisticated cellular networks. Like all engineering disciplines in the history, it's growth is hindered by the lack of, yet would be greatly benefited by, profound development of theoretical guidance.

1.2 RNA synthetic biology

There are numerous approaches to reprogram gene expression in living cells. As of today, protein mediated genetic circuits are the most studied and utilized in synthetic biology [13]. However, as next-generation sequencing technology has bloomed to uncover the mystery of RNA molecules, RNA synthetic biology has become a highly-discussed topic[53].There are several advantages of RNA regulators. First of all, RNA mediated gene regulation is simpler —it skips the translation step in the signal transduction process, saving cellular energy resources [53, 52], and thus alleviating the heavy burden synthetic circuits force on living cells. Most importantly, RNA mediated gene regulatory circuitry has the potential to be modeled very well due to its simplicity. Secondly, RNA gene regulators are very versatile. They are responsible for various cellular functions, including splicing and editing RNA, modifying rRNA, catalyzing biochemical reactions and regulating gene expression at both transcriptional and translational levels. These RNA mediated regulations are achieved by the interactions of RNA molecules with not only other RNAs, but also DNA, proteins and small molecules [53].This versatility offers more control points in gene regulation. Finally, RNA molecules degrade much faster than protein molecules. Fast degradation is a key characteristic for dynamic control of circuitry[102]. For instance, in a signal transduction cascade, RNA molecules can be briefly induced to pass a time-dependent signal pulse to the next level in a cascade; on the other hand, a protein mediated signal would remain in the on-state until all signaling proteins are slowly degraded. These advantageous characteristics of RNA regulators make them the ideal candidate parts for sophisticated genetic circuitry. One of the types of transcriptional control RNA molecules known as sRNA has been studied extensively. For example, antisense sRNA controls gene expression through transcriptional attenuation. As shown in Figure 1.1, when antisense RNA is present, it interacts

with its sense RNA target (an attenuator) to control transcription. For transcriptional repressors, this interaction causes the attenuator to form a terminator structure that terminates the transcription of the downstream coding region (Figure 1.1A) [110]. On the other hand, a small transcription activating RNA (STAR) binds to its sense target RNA to relieve the terminator structure and activate the transcription of downstream gene (Figure 1.1B)[20]. In the past few years, the Lucks Lab has created large libraries of orthogonal sRNA repressor and STAR variants derived from natural systems[20, 110, 68]. These works have cleared the barriers to building more sophisticated RNA genetic circuits.

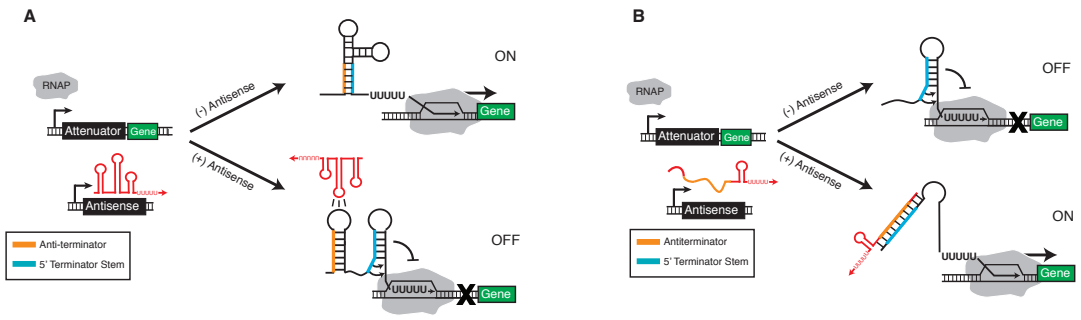


Figure 1.1: Mechanism of antisense sRNA mediated transcriptional regulation. (A) A schematic of the sRNA repressor mechanism. In the absence of antisense sRNA repressor, the anti-terminator sequence prevents the terminator formation and allows downstream transcription (ON). When present, the antisense sequesters the anti-terminator sequence and allows terminator formation, preventing downstream transcription (OFF). (B) Schematic of the STAR mechanism. In the absence of a STAR antisense, an intrinsic terminator is formed in the sense target RNA, preventing transcription elongation (OFF). In the presence of the STAR antisense, the 5' intrinsic terminator stem is sequestered by the STAR antisense, allowing downstream transcription by RNA polymerase (RNAP). Figure from Chappell et al.[20]

1.3 Mathematical framework of biological networks

Biological systems are astoundingly complex. To describe a biological network with all the detailed molecule interactions, a model would be very large and computationally expensive. To address this problem, Alon used simplified qualitative models to describe protein transcription. He showed that a handful of simple equations is sufficient to capture some essence of a complex biological system networks[1]. For instance, as shown in figure 1.3, the protein mediated transcription repression event consisting multiple biochemical reaction steps can be described with one simple Hill function containing ordinary differential equation (ODE).

$$\frac{dY}{dt} = \beta(1 - \frac{X^{*n}}{K^n + X^{*n}}) - d_Y Y \tag{1.1}$$

In this ODE in Figure 1.2, for example, β is the maximum expression level of the promoter, and K is the repression coefficient with units of concentration, which defines the concentration of active X needed to significantly repress the expression of Y. The Hill coefficient n governs the steepness of the input function, and d_Y is the degradation/dilution rate of protein Y. Our effective model is inspired by this philosophy of simplicity but applied to sRNA mediated regulation networks. The idea is that because an effective model would only contain a handful of unknown parameters, it would be possible to estimate most or even all of them. To build an effective model, a lot of the molecule integrations are coarse-grained and many parameters are unknown. Models with many unknown parameters, renormalized interactions and murky topologies are called sloppy[14]. In many cases, important predictions largely depend only on a few stiff combinations of parameters, followed by a sequence of geometrically less important sloppy ones, whose exact values do not need to be known to quantitatively understand system behavior[65]. In 2001, Bailey proposed this controversial idea that we do not need to know every unknown parameter in a model to predict a biological

system[5]. Sethna later proposed the concept of sloppy modeling and proved Baileys theory by applying sloppy models in two biochemical regulation systems in cells as test problems. Also, he showed that sensitivity analysis with the Fisher Information Matrix (FIM), which finds the most identifiable parameters within a given set of observed experimental data based on parameter sensitivities, would identify the stiff parameter combinations for model prediction[65].

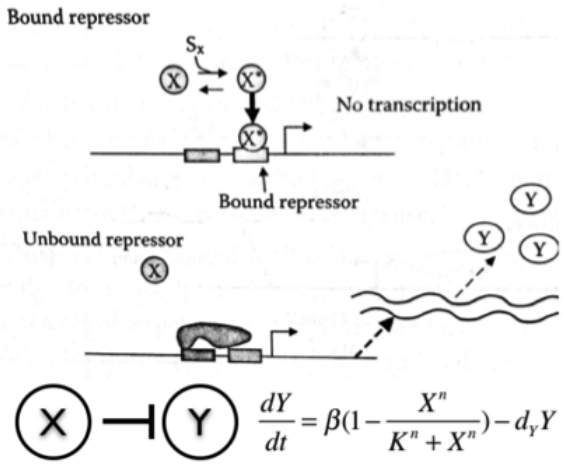


Figure 1.2: A schematic of a transcription event described by one simple Hill function. A repressor, X, is a transcription factor protein that decreases the rate of mRNA transcription when it binds to the promoter. The signal, S_x, increases the probability that X is in its active form X^*. X^* binds a specific site in the promoter of gene Y to decrease transcription and production of Y protein. One ODE is adequate to describe the entire event. Adapted from *Alon, Uri, (2006) —An Introduction to systems biology*[1].

1.4 The advantages of cell-free systems

Due to the complexities of living organisms, researchers have developed an *E.coli* based cell-free Transcription-Translation (TX-TL) system to perform basic biological processes. Cell-free reactions often consist of three components: cell extract or purified gene expression machinery, a buffer/energy mix optimized for gene expression, and DNA that encodes the genetic network. These systems do not require selec-

tion markers, so there are no limitations on plasmid origin of replication or antibiotic compatibility[106]. Also, since there is no cell membrane, DNA plasmids can be introduced to the system at any time during the reaction. These advantages allow us to rapidly prototype RNA circuitries. Takahashi et al. used this TX-TL system to characterize RNA circuitry and demonstrate its fast dynamics[108]. With the simplified model based on Alons work[1], they were able to quantitatively estimate the response time of an RNA signal transduction cascade to be five minutes per step of the cascade[108]. Takahashis work was the launching point of the first part of this work, which is to quantitatively model the same cascade in TX-TL systems and identify all unknown parameters. Fortunately, it is trivial to obtain experimental data for model training because of the advantages of TX-TL systems.

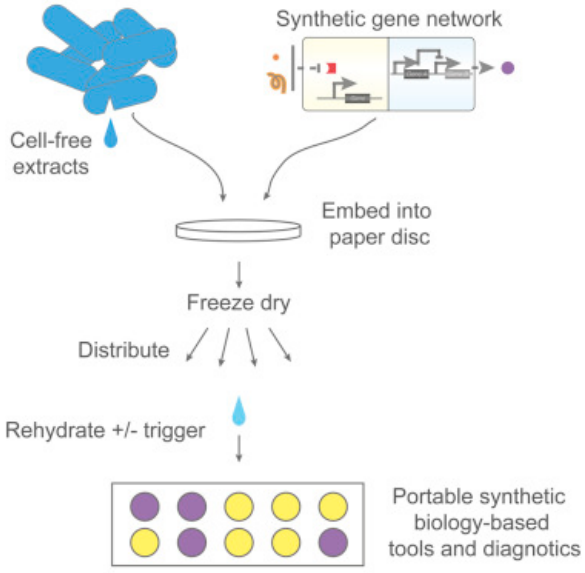


Figure 1.3: Schematic describing the process of arraying synthetic gene networks on paper using printed arrays. Figure from *Pardee et al* [84].

TX-TL is also a particularly ideal environment for model development. A coarse-grained effective model is built on numerous assumptions. For instance, in Equation 1.1, we assume the rate of transcription and the rate of degradation are constant and the

transcriptional factor is abundant. These assumptions often ignore resource reallocation, stress response ,noise propagation and more complex events in the circuit environment. Therefore might not be sufficient to capture the exact dynamic of a synthetic circuit. TX-TL, on the other hand, is a much simpler single-cell-like environment where these assumptions could hold. Models based on TX-TL experiments provide more accurate information on the synthetic circuit and offers more insights to develop a reliable design principle for RNA circuitry. TX-TL also has a powerful diagnostic application potential. Recently, a new method of embedding cell-free synthetic gene networks onto paper and other materials was developed by the Collins Lab[84]. This work showed that cell-free systems can be freeze dried, then preserved and stored in room temperature on paper. The embedded material can then be activated by simply adding water. This drastically reduces the cost of storage and transportation of cell-free systems. As a sensing tool, they showed that it could be set up in a way to produce a color pigment output that is visible to naked eyes in response to a certain inducer (Figure 1.3)[84]. Therefore, the sensor could be implemented in a low-tech environment and operated by people without special training. This method opened up a great number of possible applications for synthetic biology.

1.5 Building theoretical foundation for RNA synthetic biology

1.5.1 Chapter 2: Building the first predictive model for RNA circuitry

Fast dynamic is one of the major advantages of regulatory sRNA. It empowers synthetic biology and adds sophistication into synthetic circuits. In order to fully utilize this ad-

vantage, accurate prediction and full control of dynamic of RNA circuits were essential. However, little work has been done to model RNA circuits. To clear this roadblock, the first part of the thesis we set off to test whether we can build an effective model to describe the complex RNA regulation mechanism. We started building the effective model for a 3 level repression sRNA cascade with the minimal numbers of ordinary differential equations (ODEs) base on qualitative experimental observations in TX-TL. In order to obtain accurate predictions of these circuits, the unknown parameters need to be estimated. We then used a sensitivity analysis based parameterization method to estimate all involved parameters using dynamic performance of circuits that are composed of parts of the cascade in TX-TL. This method yielded distributions of all 13 parameters. We demonstrated the accuracy of these parameters by testing them with new construct dynamic predictions. We found that these predictions all agreed with experimental data quantitatively. Finally, we used the method to investigating the well documented but little understood batch-to-batch difference in TX-TL systems. We parameterized two sets of parameters from two different batches and examined their differences and similarities. We found that the parameters that differed the most between two batches were all energy resource related. This part of the thesis established a fundamental method of modeling and obtaining parameters for sRNA regulated circuits.

1.5.2 Chapter 3: Apply the predictive model to design and engineer complex RNA circuitry

Next we expand the repertoire of synthetic gene networks built from these regulators by constructing a transcriptional negative autoregulation (NAR) network out of small RNAs (sRNAs). NAR network motifs are core motifs of natural genetic networks, and

are known for reducing network response time and steady state signal. Here we use cell-free TX-TL reactions and a computational model to design and prototype sRNA NAR constructs. In our *in silico* design phase, we find that due to the weak repression strength and fast degradation rate of sRNA molecules, two pieces of sRNA need to be stacked in tandem to achieve the functionality of a NAR circuit. In order to engineer this design successfully, we shortened the length of sRNA and terminator pieces significantly while preserving the repressor's functionality. Using parameter sensitivity analysis, we design a simple set of experiments that allow us to accurately predict NAR dynamic in TX-TL. Both of our predictive model and the TX-TL prototyping experiments confirm the necessity of a double repressor design. Finally, We transfer successful network designs into *Escherichia coli* and show that our sRNA transcriptional NAR network reduces both network response time and steady-state gene expression, thus confirmed the NAR network's functionality. This work broadens our ability to construct increasingly sophisticated RNA genetic networks with predictable function.

1.5.3 Chapter 4: Define the design principle for transfer function manipulation with sRNA

Inducible promoter systems are at the heart of many applications in molecular biology, allowing researchers to control gene expression by addition of small molecules. While there is a range of potentially useful inducible promoter systems, their transfer functions—the relationship between molecular input and molecular output—are static and often require tuning via molecular engineering. In previous work, we showed that Small Transcription Activating RNAs (STARs) can serve as an intermediate between *Lux* operon and reporter gene GFP to form a function STAR-hybrid operon. Interest-

ingly, we observed three distinct transfer functions with three different STARs. In this part of the work, we started with a theoretical approach to understand what parameters contribute to this transfer function shift. We then estimated these parameters with a sensitivity analysis based parameterization method. These estimated parameters were later used to generate simulated transfer functions that showed close resemblance with our experimental result. With a theoretical approach, we found that this configuration of STAR hybrid operons offers an tuning knob at the transcriptional level with a insulated, interchangeable constitutive promoter. We used a *Xyl* operon to showcase this application experimentally and demonstrated our theory in practice. Later we found that this configuration also offers a dynamic rage (ON/OFF state) amplification across all three pairs of STARs. We discovered that this phenomenon is caused by a slowdown in RNA degradation at the on state due to saturation of RNA degradation machineries. Our simulation also offered a design guideline for this amplification effect to take place. Next, we used STAR sequesters to further improve the dynamic range of a STAR hybrid *Xyl* operon. Finally, we showed that this engineering approach has a broad application by showing successful dynamic range amplification of a *Lac* STAR operon and a STAR hybrid riboswitch. We ended with a discussion on limitation and cellular environment context dependence by showing the dynamic range amplification performances in various *E.coli* strains. We hope that this work will improve our understanding and the ability to control biological systems.

CHAPTER 2

GENERATING EFFECTIVE MODELS AND PARAMETERS FOR RNA GENETIC CIRCUITS[1]

2.1 Abstract

RNA genetic circuitry is emerging as a powerful tool to control gene expression. However, little work has been done to create a theoretical foundation for RNA circuit design. A prerequisite to this is a quantitative modeling framework that accurately describes the dynamics of RNA circuits. In this work, we develop an ordinary differential equation model of transcriptional RNA genetic circuitry, using an RNA cascade as a test case. We show that parameter sensitivity analysis can be used to design a set of four simple experiments that can be performed in parallel using rapid cell-free transcription-translation (TX-TL) reactions to determine the thirteen parameters of the model. The resulting model accurately recapitulates the dynamic behavior of the cascade, and can be easily extended to predict the function of new cascade variants that utilize new elements with limited additional characterization experiments. Interestingly, we show that inconsistencies between model predictions and experiments led to the model-guided discovery of a previously unknown maturation step required for RNA regulator function. We also determine circuit parameters in two different batches of TX-TL, and show that batch-to-batch variation can be attributed to differences in parameters that are directly related to the concentrations of core gene expression machinery. We anticipate the RNA circuit models developed here will inform the creation of computer aided genetic circuit

[1]THIS WORK WAS ORIGINALLY PUBLISHED IN ACS SYNTHETIC BIOLOGY AND HAS BEEN REPRODUCED HERE WITH PERMISSION. HU, C. Y., VARNER, J. D., & LUCKS, J. B. (2015). GENERATING EFFECTIVE MODELS AND PARAMETERS FOR RNA GENETIC CIRCUITS. ACS SYNTHETIC BIOLOGY, 4(8), 914926. HTTP://DOI.ORG/10.1021/ACSSYNBIO.5B00077

design tools that can incorporate the growing number of RNA regulators, and that the parameterization method will find use in determining functional parameters of a broad array of natural and synthetic regulatory systems.

2.2 Introduction

A central goal in synthetic biology is to engineer biological systems to optimally perform natural and sometimes novel tasks. These tasks are varied, and are often related to important challenges in sustainability and health. For example, there has been a great deal of effort placed on harnessing the natural capabilities of cells to synthesize complex molecules from renewable feedstocks, and to sense and respond to dynamically changing environments. These capabilities are themselves implemented through genetic circuits networks of interacting gene expression regulators that can dynamically balance metabolic pathways [131] and integrate multiple signals to make behavioral decisions[61], among many other behaviors related to cell replication and survival[114]. Thus a cells behavior is intimately connected to the structure of its genetic circuits[131, 70], making the precise engineering of genetic regulators and networks a central goal of synthetic biology [61, 87].

Over the past decade, synthetic biology has seen significant advances in the quantity and sophistication of engineered genetic circuits[114, 76]. Recently, regulatory RNAs have emerged as powerful components of the synthetic biology toolbox for constructing genetic circuits that control the timing and amount of gene expression[131, 21]. These natural and synthetic RNAs are diverse, and can regulate transcription, translation and RNA degradation, especially in bacteria[61, 88]. There is also a growing body of work on developing design principles for engineering the functional properties of reg-

ulatory RNAs. This work makes use of a powerful suite of computational[114, 130] and experimental[131, 63, 70] methods to discern the underlying RNA structures behind the regulatory function[61, 19, 87], and is leading to new classes of highly-designable RNA regulators[114, 40, 76].

Of particular interest are RNA mechanisms that regulate transcription. These regulators are special because they can be wired together into RNA-only genetic circuits that propagate information as RNAs without the need to translate or degrade intermediate regulatory proteins[64]. Recently it has been shown that RNA transcriptional repressors, also known as attenuators[12] can be configured into NOR logic gates and transcriptional cascades[64]. They can also be used to construct more sophisticated circuits, such as single-input modules[90] that sequentially activate the expression of multiple target genes[108]. Advances in RNA engineering approaches have also greatly expanded the types of circuits that can be built out of RNA mechanisms. For example, the creation of small transcription activating RNAs, or STARs, allows the creation of new types of RNA logic gates that implement AND and NIMPLY[20]. This, combined with exciting new developments in using the RNA-protein hybrid Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR) interference (CRISPRi) system to construct transcriptional repressors[38], activators[9], and NOT and NOR logic gates in cells[77], has started to draw attention to RNA-based genetic circuits as a platform for precisely controlling gene expression.

A major gap in our RNA regulatory toolbox is the lack of a computational framework that can be used to model and ultimately design RNA genetic circuitry. Such a modeling framework is also necessary for incorporating RNA regulators and circuits into a growing suite of computer-aided design (CAD) tools that allow users to use high-level cellular behavioral specifications to design synthetic genetic circuits and select

genetic regulatory parts that implement those behaviors[26, 18, 58]. While there has been progress in modeling the impact of individual RNA regulators on tuning gene expression[16, 8, 102], there has been little work in modeling how networks of transcriptional RNA regulators impact the coordinated expression of multiple genes. In contrast, for protein-based genetic circuits, systems of ordinary differential equations (ODEs) that model the basic processes of gene expression in a genetic network as coupled chemical reactions[72] are commonly used to computationally study both natural[1] and synthetic[82] genetic networks. However, it is generally not known if simple ODE-based frameworks work for modeling RNA transcriptional circuits, and if they do, what parameter values are needed for accurate prediction of circuit behavior (Figure 1.1A).

In this work, we develop an effective modeling framework that quantitatively captures the dynamic outputs of RNA-only transcriptional circuits. To do this, we leverage recent advances in using cell-free transcription-translation (TX-TL) reactions as rapid genetic circuit prototyping environments for characterizing genetic circuit dynamics[79, 19, 108] and for modeling gene expression processes[57]. We build this model by studying a double-repression RNA-only transcriptional cascade[64, 108]. We start by systematically constructing a system of ODEs that model the expression and degradation of each RNA regulator in the cascade (Figure 2.1A), and find that RNA maturation delay equations must be added in order to qualitatively capture cascade function. Since many of the model parameters for RNA-only circuits are not available in the literature, we next adopt a systems biology approach to estimate all unknown parameters in this model. Rather than perform a suite of biochemical experiments to measure each parameter in turn, this approach uses sensitivity analysis on the mathematical structure of the genetic network to design a minimal set of characterization experiments that can be used to rapidly and quantitatively determine all parameters in a model[35] (Figure 2.1B). Specifically, we show that all thirteen parameters of our model can be estimated

from only four TX-TL experiments that can be run in parallel in under two hours. We then use our estimated parameters with the governing ODE framework to predict the function of six new RNA transcriptional circuit variants and show that the simulated predictions compare favorably with experimental measurements. Finally, we perform model parameterization experiments using a different batch of TX-TL reagents, and show that batch-to-batch variation can be attributed to differences in parameters that are directly related to the concentrations of core gene expression machinery. We end with a discussion about how this method can be generalized to rapidly determine parameters for quantitatively modeling the dynamics of synthetic genetic circuits.

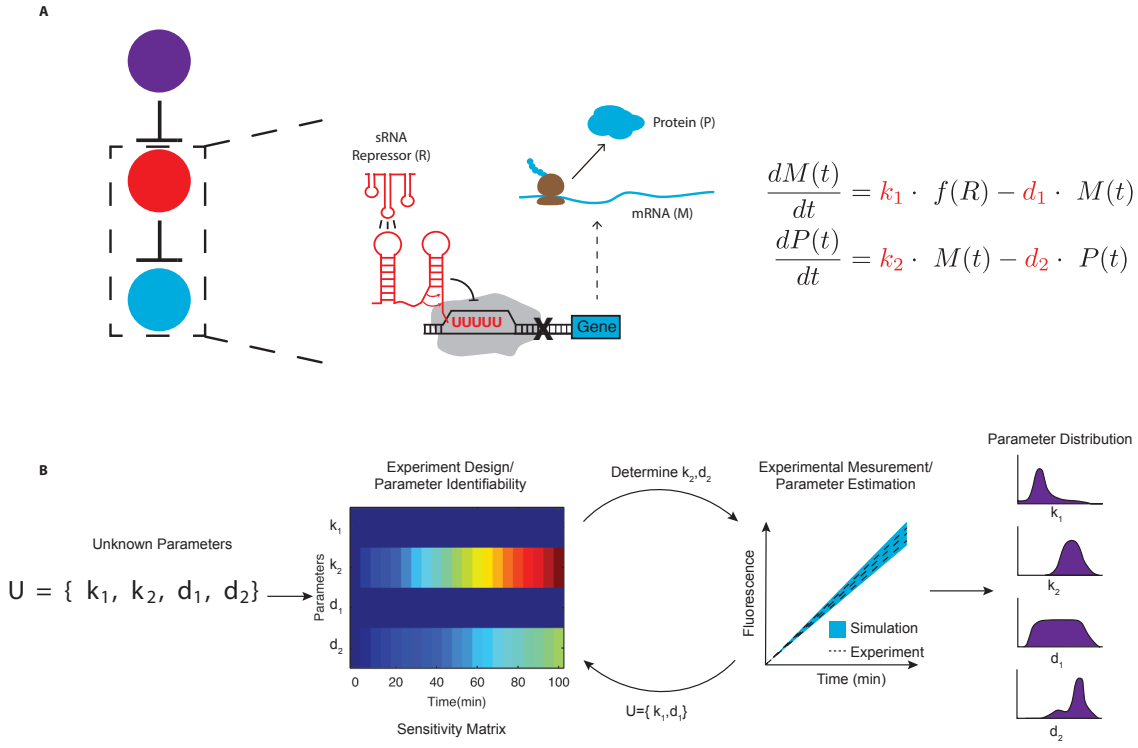


Figure 2.1: Schematic of the model development and parameterization process. (A) A 3-level, DNA-encoded RNA transcriptional cascade (colored circles) is composed of two orthogonal sRNA repressor/attenuator pairs. The RNAs are configured in a double repression cascade, with the final output being the transcription of a target gene that encodes a translated protein. The complex mechanism of RNA transcriptional repression can be described by coarse-grained ordinary differential equations (ODEs) with a handful of unknown parameters (red). (B) Parameterization experiments can be designed based on a parameter sensitivity analysis of the model equations. This analysis identifies which parameters can be estimated from a particular experimental design. These experiments can then be performed to estimate the indicated parameters. This process can then be iterated until all parameters are estimated, resulting in distributions of parameters that accurately model the desired genetic circuitry.

2.3 Results and Discussion

2.3.1 Model Derivation

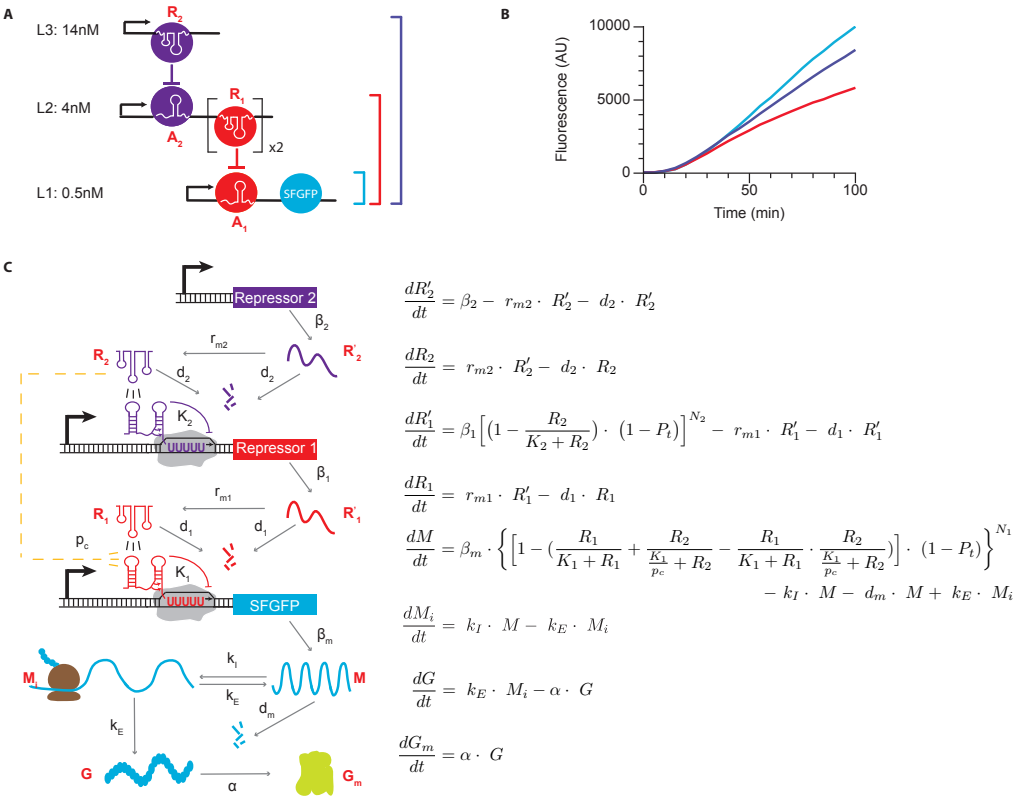
Our first goal was to develop a modeling framework that could qualitatively capture the dynamical behavior of RNA-only circuits. One mathematical framework that has found wide use in modeling genetic circuits is systems of ordinary differential equations that treat the basic processes of gene expression in a genetic circuit as coupled chemical reactions[1]. While these models can vary in detail, in essence they consider the concentration of a given molecular species in time to be a function of its synthesis and degradation rates[1] (Figure 2.1A). These rates in turn can be functions of the concentrations of other regulators in the circuit, effectively coupling the equations according to the circuits network topology.

Our goal was to construct the simplest possible model (i.e. fewest number of equations and parameters) that could capture the function of RNA transcriptional repressors and cascades. To do this, we focused on modeling a two-repressor transcriptional cascade that had been previously characterized[64, 108]. This cascade was constructed from two orthogonal RNA repressors called transcriptional attenuators (Figure 2.2A). Transcriptional attenuators act as genetic switches by blocking or allowing transcription through conditional formation of an RNA intrinsic terminator hairpin[81]. In the absence of an antisense sRNA, sequences in the attenuator prevent terminator formation, which allow transcription of a downstream gene. When antisense sRNAs are present, they bind to their attenuator targets, which allows terminator formation and switches transcription to an OFF state.

The double-repression RNA transcriptional cascade consisted of three levels[64,

108] (Figure 2.2A). The bottom level of the cascade (L1) consisted of a constitutive promoter followed by attenuator-1 (A1), which controlled the transcription of a downstream super folder GFP (SFGFP)[86] coding sequence. A1 was itself switched to an OFF state by interactions with repressor sRNA 1 (R1), which was present in two tandem copies on the second level (L2) of the cascade[64, 108]. The complete L2 also contained a constitutive promoter followed by attenuator-2 (A2), both upstream of the R1 copies. Following previous work, self-cleaving ribozymes were included before each R1 copy to ensure proper function[64]. This configuration allowed the transcription of R1 to be controlled by repressor sRNA 2 (R2), which was expressed from a constitutive promoter on level 3 (L3) of the cascade.

Characterization of the cascade in TX-TL reactions revealed the patterns of fluorescence expected from combining different levels of the cascade in the reactions (Figure 2.2B). TX-TL reactions consist of three components: cell extract, energy solution/buffer and input circuit DNA. To characterize the performance of RNA genetic circuits, plasmid DNA encoding different combinations of cascade levels were mixed with extract and buffer following previously published protocols[106] (see Methods). These were then monitored on a plate reader to measure SFGFP fluorescence over time to characterize overall circuit expression. As expected, when only L1 DNA was present, we observed a rapid increase in SFGFP fluorescence, which was decreased when both L1 and L2 DNA were present (Figure 2.2B). The addition of L3 DNA in the reaction showed an increase over the L1+L2 condition indicating that the double repression of the cascade was functioning properly as had been observed previously[108]. As a starting point for our model, we considered the expression of each RNA in the system to be governed by one mass balance ODE that accounts for its synthesis and degradation rates. For simplicity, we considered these rates to be described by single constants. To model transcriptional repression, we introduced Hill functions of order one, based on previ-



$$\frac{dR_2'}{dt} = \beta_2 - r_{m2} \cdot R_2' - d_2 \cdot R_2'$$

$$\frac{dR_2}{dt} = r_{m2} \cdot R_2' - d_2 \cdot R_2$$

$$\frac{dR_1'}{dt} = \beta_1 \left[\left(1 - \frac{R_2}{K_2 + R_2}\right) \cdot \left(1 - P_t\right) \right]^{N_2} - r_{m1} \cdot R_1' - d_1 \cdot R_1'$$

$$\frac{dR_1}{dt} = r_{m1} \cdot R_1' - d_1 \cdot R_1$$

$$\frac{dM}{dt} = \beta_m \cdot \left\{ \left[1 - \left(\frac{R_1}{K_1 + R_1} + \frac{R_2}{\frac{K_1}{p_c} + R_2} - \frac{R_1}{K_1 + R_1} \cdot \frac{R_2}{\frac{K_1}{p_c} + R_2} \right) \right] \cdot \left(1 - P_t\right) \right\}^{N_1} - k_I \cdot M - d_m \cdot M + k_E \cdot M_i$$

$$\frac{dM_i}{dt} = k_I \cdot M - k_E \cdot M_i$$

$$\frac{dG}{dt} = k_E \cdot M_i - \alpha \cdot G$$

$$\frac{dG_m}{dt} = \alpha \cdot G$$

Figure 2.2: Governing Equations for a 3-level sRNA transcriptional cascade (A) Schematic of the cascade. Level 1 (L1) plasmid expresses attenuator 1 (A_1) controlling SFGFP transcription; Level 2 (L2) expresses attenuator 2 (A_2) controlling the transcription of tandem copies of sRNA repressor 1 (R_1); Level 3 (L3) expresses sRNA repressor 2 (R_1). Concentrations of DNA templates used in TX-TL reactions for part B are listed beside the levels. (B) Representative fluorescence signal time trajectories in TX-TL reactions containing three different combinations of DNA templates from the transcriptional cascade in part A. L1 alone leads to a high rate of SFGFP production (blue line); L1+L2 results in results in a reduced SFGFP production rate (red line) due to R_1 repressing A1; L1+L2+L3 (purple line) results in a higher SFGFP production rate than just L1+L2 due to the double negative inversion of the full cascade. (C) Schematic (left) of the mechanistic steps of the cascade in part A that are captured by the governing equations (right). The equations model the tandem copies of R_1 on L2 as one repressor.

ous experimental work in characterizing sRNA/attenuator transfer functions[64]. This allowed us to construct a set of four equations with eight parameters that captured the flow of information in the RNA transcriptional cascade (Appendix A Equations A.1-A.4, Figure A.1.1). Note that we incorporated the tandem copies of R1 present in L2 of the cascade through the parameters of the model rather than extra mechanistic steps.

Using this simple model, we simulated time dependent trajectories of the cascade (Figure A.2). These results showed that there were several qualitative disagreements between the model and the previously reported experimental data for this cascade. In particular, several mechanistic details of RNA transcriptional attenuators were not included in these equations. The most important of these is cross-talk, or the ability for non-designed interactions to cause repression between different levels of the cascade due to the imperfect orthogonality of the repressors used in the system[64]. We incorporated cross-talk into the model by using a constrained fussy logic formulation, treating the bottom level of the cascade as an OR gate module with two signaling species accounting for the contributions of cognate and cross-talk interactions[71](Figure A.1B,Equations A.5-A.8).

Another feature of the RNA transcriptional attenuators not captured is their ability to be placed in tandem next to each other so that the combined attenuator can respond to multiple antisense sRNAs. Previous work had found that attenuators in tandem multiplied their effects and increased their sensitivity to antisense RNA[64]. In fact this feature was exploited in order to create a time-delay in a singe-input module that could sequentially activate the expression of two different genes[108]. To incorporate tandem attenuators in our model, we raised the repressive Hill functions to the power of tandem attenuator number on that level (Figure A.1C, A.2B, A.2 Equations A.9-A.12). Previous work also showed that some aspect of the attenuation mechanism causes repression of

the downstream gene even in the absence of any antisense sRNA[64], which was also incorporated into our model (Figure A.1D, A.2C, Equations A.13-A.16).

Table 2.1: Model Parameters

Model Parameters	Num.	Definition	Unit
β_2	1	Maximal transcriptional rate of level of R_2 (determined from β_m)	Conc./sec
β_1	2	Maximal transcriptional rate of level of R_1 (determined from β_m)	Conc./sec
β_m	3	Maximal transcriptional rate of M	Conc./sec
K_2	4	Repression coefficient of R_2 on the production of R1	Conc.
K_1	5	Repression coefficient of R_1 on the production of M	Conc.
k_E	6	Translational elongation rate of SFGFP	1/sec
d_2	7	Degradation rate of the repressor R_2	1/sec
d_1	8	Degradation rate of the repressor R_1	1/sec
d_m	9	Degradation rate of M	1/sec
α	10	Maturation rate of SFGFP protein	1/sec
k_I	11	Translational initiation rate of SFGFP	1/sec
r_{m2}	12	Maturation rate of R_2	1/sec
r_{m1}	13	Maturation rate of R_1	1/sec
p_c	14	Relative crosstalk level 1 experiences from level 3	N/A
P_t	15	Probability of that an attenuator auto terminates itself	N/A
N_1	16	Number of tandem attenuator in level 1	N/A
N_2	17	Number of tandem attenuator in level 2	N/A

We next began comparing our model to measured fluorescence trajectories from TX-TL reactions (Figure A.3). Since we used a fluorescent reporter protein as a final output of the cascade, we added additional equations to model the translation and maturation of

Table 2.2: Model Species

Model species	Definition	Unit
$R_2^{'}$	Immature repressor R_2	Conc.
R_2	Mature repressor R_2	Conc.
$R_2^{'}$	Immature repressor R_1	Conc.
R_1	Mature repressor R_1	Conc.
M	mRNA of SFGFP protein	Conc.
M_i	Translationally Initialized mRNA of SFGFP protein	Conc.
G	Immature SFGFP protein	Conc.
G_m	Observable SFGFP protein	Conc.

the fluorescent reporter protein (Equations A.17-A.22). Specifically, we modeled translation as a 2-step process consisting of initiation and elongation, and ignored SFGFP degradation which is appropriate for TX-TL reactions[106]. During this model formulation process, we noticed that there was a delay between when DNA was introduced in the TX-TL reactions, and the time it took to observe fluorescence that was longer than the expected delay due to SFGFP maturation. Specifically, we were not able to observe a fluorescence signal until 20 minutes after the reaction was initiated. As we tried to qualitatively fit the model to the experimental trajectories, we found that this delay was too large to be described by any known mechanism of this system (Figure A.3A, A.3B). Thus we hypothesized that the observed delay was specific to the TX-TL system, which could be due to the time needed for activation of extract core machinery after mixing with buffer. To test this, we pre-incubated extract and buffer together at 37°C for 20 minutes before adding DNA and performing the circuit characterization. We found that when extract and buffer were pre-incubated, this delay was removed, allowing greater qualitative agreement with our model (Figure A.3C, A.3D).

While we were able to obtain a qualitative match between our model and experimental trajectories, we noticed that when L1 and L2 were both present in TX-TL reactions, we observed a decrease in SFGFP production rate after 30 minutes. This manifested as a downward bending of the L1+L2 fluorescence trajectories that we could not qualitatively capture (Figure A.4). We hypothesized this was due to the fact that the sRNA R1 repressor encoded in the L2 plasmid did not function immediately once synthesized, which was an underlying assumption of the model up to this point. To incorporate this into our model, we added additional sRNA maturation steps into the governing equations (Equations A.23-A.30, Figure 2.2C). As shown in Figure A.2B, after we introduced sRNA maturation delay terms into the governing equations, we were able to qualitatively capture the bending of the L1+L2 trajectories.

Since the same promoter was used on all constructs, β_1 and β_2 were determined by multiplying β_m by an appropriate factor based on DNA template concentrations. The final governing equation set thus consisted of eight ODEs and thirteen unknown parameters (Figure 2.2C,Tables 2.1 and 2.2), which qualitatively captured the behavior of the three-level RNA transcriptional cascade.

2.3.2 Parameter Estimation Through Sensitivity Analysis-Based Experimental Design

Our next goal was to determine parameters that would give quantitative agreement between our model and measurements of the final output fluorescence of the RNA transcriptional cascade. Parameter estimation is a highly non-trivial problem, as parameters are complex functions themselves of many detailed biochemical reactions. For example, transcription rate is commonly modeled as a single parameter with units

of nucleotides/sec, when in reality it encompasses many separate processes including polymerase-promoter recognition, open-complex formation, escape probability and non-uniform elongation rates[92]. The most common method for determining parameters is to find them in the literature, and optionally consider parameter ranges centered on literature values[16]. While useful, this approach has limitations, including the fact that literature parameters are often not measured in an experimental/cellular/construct context relevant to the functioning of the model circuit, parameter measurement may not be consistent with the approximations made by the modeling framework, and it may be impossible to find certain parameters especially in cases when a synthetic regulator does not exist in nature[20]. An alternative approach is to perform a series of specific experiments designed to isolate the measurement of each parameter. While effective[83], this approach is difficult to scale as the size of the genetic networks grow.

Potentially more powerful are methods that can take into account the structure of the genetic circuit to design a minimal set of experiments that can be used to rapidly and quantitatively determine all parameters in a model. Such methods have been developed in the context of systems biology, which focuses on understanding a biological systems structure and dynamics. Because natural biological networks are usually massive and full of unknown species and parameters, systems biologists have developed numerous methods to locate the most important components of a genetic network and identify the most sensitive parameters[61]. Here we adapted a technique based on parameter sensitivity analysis and used it to design experiments that would provide enough information to identify all unknown parameters in our RNA circuitry model (Figure 2.1).

The basis of this technique is to use sensitivity analysis to guide the design of experiments that can use the full time trajectory information of a genetic circuits output to determine multiple parameters at the same time (Figure 2.1). For a particular ex-

periment modeled by a system of ODEs, sensitivity analysis is based on the sensitivity matrix, which describes how the time varying molecular concentrations of the genetic circuit in the experiment change in response to a change in the parameters of the model (see Methods)[42]. Since our experiments measure SFGFP expression, we focused our analysis on how sensitive predicted trajectories of SFGFP expression were to changes in the thirteen unknown parameters. Parameters with large magnitudes in the sensitivity matrix (highly sensitive) can then be determined by fitting them to make the model match the experimental data. By proposing different experiments, this procedure can be iterated until a panel of experiments is designed that together can be used to estimate all parameters of the model (Figure 2.1). We note that this procedure is particularly amenable to being used with TX-TL reactions since circuit DNA template concentrations can be easily varied to rapidly design a set of parameterization experiments.

To perform this procedure, we used an initial set of parameter guesses taken from the literature or manual fitting to calculate the sensitivity matrix from a proposed experimental design. Our goal was to determine a reduced set of TX-TL experiments that could be used to find all thirteen parameters of our RNA transcriptional cascade model. In order to strike a balance between TX-TL energy resource usage and potential bleaching effects of the fluorescence measurement, we targeted experiments that could be performed at 29°C for 100 minutes with fluorescence collection every 5 minutes. We first performed the sensitivity analysis on an experiment consisting of just the bottom level (L1) of the cascade (Figure 2.3A). The calculated sensitivity matrix for the subset of equations that model L1 revealed that the SFGFP fluorescence trajectory in this experiment is most sensitive to the parameters β_m (promoter strength), k_E (translational elongation rate), d_m (mRNA degradation rate), α (SFGFP maturation rate), and k_I (translational initiation rate) (Figure 2.3A). This recapitulates the relationship that steady-state SFGFP concentration should be related to the product of the transcription and translation rates divided

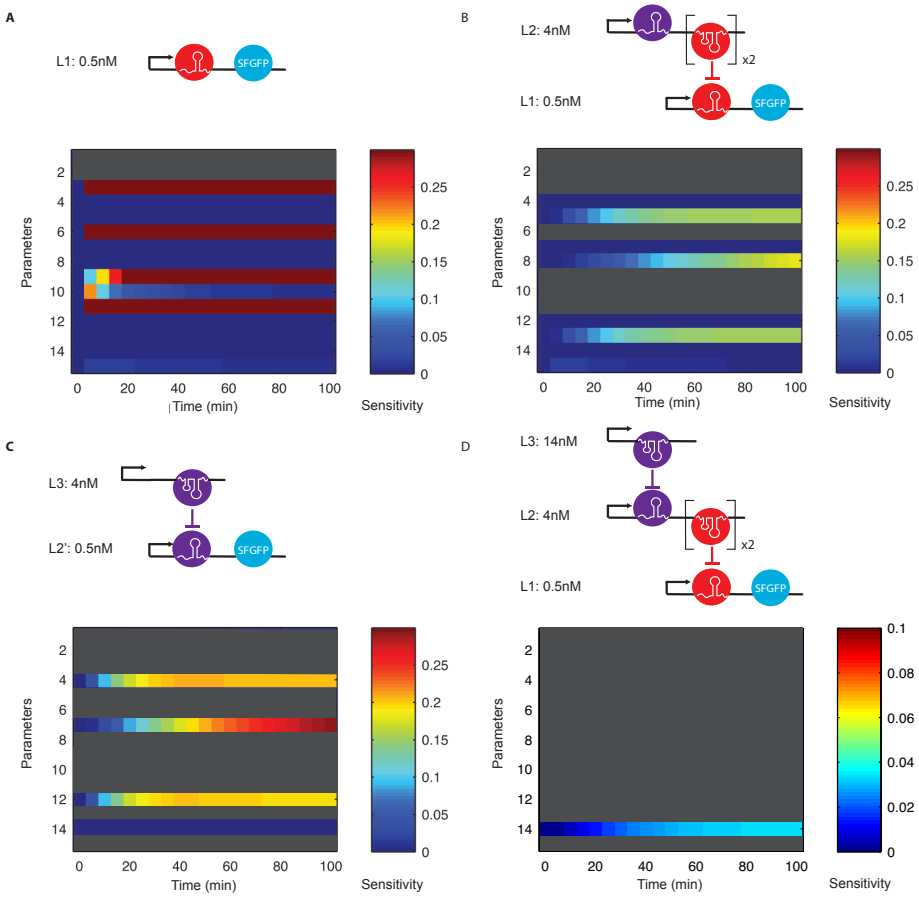


Figure 2.3: Sensitivity matrices for parameter identification experiments. Four parameterization experiments were designed based on sensitivity analysis to estimate all thirteen unknown parameters in our model (Figure 2.2). For each experiment, the constructs used are shown above the SFGFP portion of the calculated sensitivity matrix for that experiment. Parameters are numbered according to Table 2.1. Red/blue indicates high/low sensitivity, respectively. Note that time varying changes in parameter sensitivity indicates portions of the trajectories that are influenced by each parameter. Experiments were designed in order from (A) to (D), with parameters identified by previous experiments marked as grey rows. Since the same promoter was used on all constructs, β_1 and β_2, determined by multiplying β_m by an appropriate factor based on DNA template concentrations, are absent from the sensitivity analysis. (A) 0.5nM of L1 plasmid alone is able to identify five parameters: β_1, k_E, d_m, α and k_I. (B) 0.5nM of L1 and 4nM of L2 is able to identify four parameters: K_1, d_1, r_{m1} and P_t. (C) 0.5nM of L2 (Table A.2) and 4nM of L3 is able to identify three parameters: K_2, d_2, r_{m2} (D) 0.5nM of L1, 4nM of L2, and 14nM of L3 is able to identify the last parameter, p_c. The sensitivity color scale was changed in D to aid in visualization.

by the degradation rates of mRNA and SFGFP.

We next proposed an experiment that added a single layer of repression (Figure 2.3B). Since β_m, k_E, d_m, α and k_I were already identified in the previous experiment, their rows in the SSM were all set to 0 so that the algorithm skipped identified parameters from its previous rounds and searched for the next most identifiable parameters. This second experiment was additionally able to identify 4 more parameters (K_1, d_1, r_{m1}, P_t) (Figure 2.3B). Successive iterations allowed us to find two additional experiments that allowed all thirteen parameters to be identified with a total of four TX-TL experiments (Figure 2.3C,D).

We next sought to use this designed set of experiments to estimate the parameters in our model. We first performed replicate TX-TL experiments with each of the plasmid combinations we designed, and collected the fluorescence time trajectories of the reactions. Using this experimental data, we then estimated parameters using an iterative fitting procedure that used each experiment in turn to find its designated parameters (see Methods). Rather than focus on a single set of optimal parameters, we considered variations of parameter values that can capture the natural variation in experimental conditions. To do this, we initially input sets of parameters, drawn from uniform distributions centered around an initial best guess, into the fitting procedure and optimized the values of each parameter within each input set (see Methods). This resulted in 10,000 sets of the thirteen estimated parameters, which allowed us to calculate mean predicted trajectories with 95% confidence intervals (Figure 2.4).

As mentioned before, there were two R_1 repressors encoded in L2 and one R_2 repressor in L3 of the cascade, though they were both treated as single repressors in the model. Because the model was not given explicit information about the difference between these configurations, this gave us the opportunity to examine the estimated parameters to test

the adaptability of our parameter estimation procedure to this type of model discovery (Table A.4). Several estimated parameters showed that indeed this was the case. In particular, the repression coefficient of R_1 (K_1, mean 240.9) was significantly greater than the repression coefficient of R_2 (K_2, mean 132.0), indicating a weaker repression made by the double repressor configuration. This was actually consistent with previous *in vivo* characterization experiments, which showed that the L2 configuration containing an attenuator followed by tandem ribozyme-antisense coding regions was less efficient at repression than a single bare antisense[64]. In addition, the degradation rate of R_1 (d_1, mean $0.83x10^{-3}s^{-1}$) was also noticeably smaller than the degradation rate of R_2 (d_2, $2.08x10^{-3}\ s^{-1}$), showing that it is slower to degrade two repressors than it is to degrade one. Finally, the maturation rate of R_1 (rm1, mean $3.21x10^{-5}s^{-1}$) was smaller than the maturation rate of R_2 (r_{m2}, $1.55x10^{-4}s^{-1}$), indicating additional processing steps are needed for the tandem repressor configuration. This makes intuitive sense given that ribozymes were included between the tandem R1 units, which must fold and cleave before R_1 can properly function[64].

To validate our method, we compared measured trajectories of SFGFP fluorescence for each experiment to simulated trajectories using 1,000 sets of randomly drawn parameters from the 10,000 parameter set distributions (Figure 2.4). This allowed us to compare the experiments to the mean simulated trajectory and 95% confidence intervals, which captured the range of trajectories predicted by our parameter distributions. As shown in Figure 2.4, each mean simulated trajectory accurately described the experimental observations, with almost every measured experimental trajectory lying within the simulation 95% confidence intervals.

These results validated our hypothesis that sRNA genetic circuits could be accurately modeled using simplified sets of ODEs with appropriate parameters. It also showed that

our effective ODE model was able to quantitatively capture the complex biochemical conversions that take place within our model RNA circuit. Finally, it proved that our approach of estimating unknown parameters, using a minimal set of experiments designed from sensitivity analysis, was efficient, and produced accurate results. Conveniently these four TX-TL experiments could all be performed at the same time, allowing the simultaneous determination of all parameters in our model in a matter of hours. This approach is also highly generalizable to other ODE systems, and could be used to parameterize a broad array of synthetic genetic circuits.

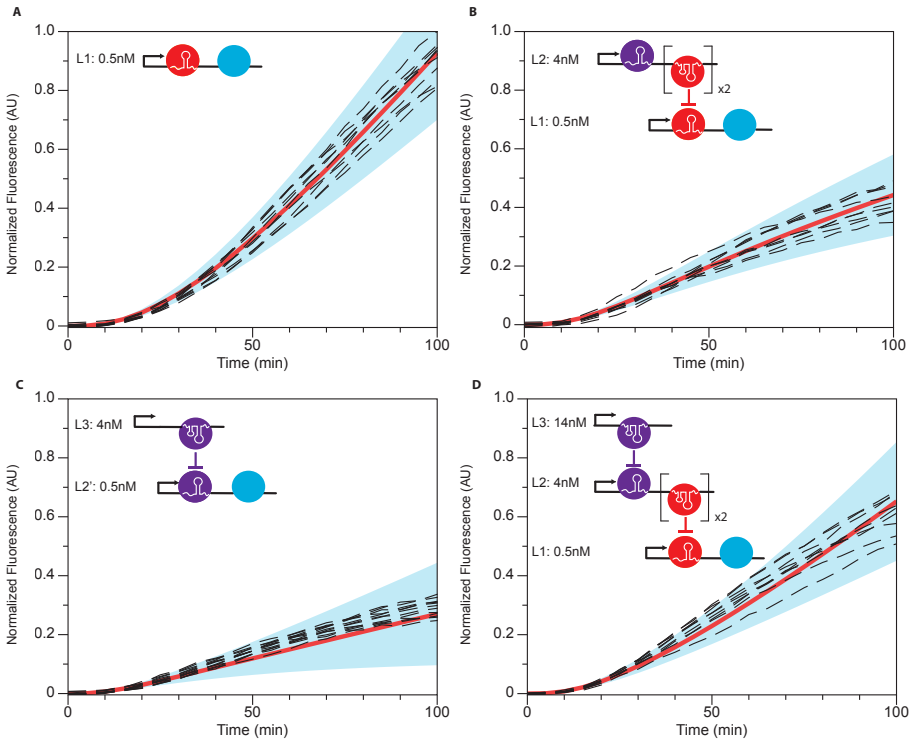


Figure 2.4: Validation of model simulations of parameter estimation experiments. Comparison of experimental trajectories of SFGFP fluorescence in TX-TL experiments (black dash lines) with simulated model predictions. Model simulated trajectories were generated by performing 1,000 simulations with parameters drawn from the set of 10,000 determined from the estimation procedure (see Methods). Experimental and model trajectories were normalized by the maximum observed experimental fluorescence (see Methods). The mean simulated trajectory (red line) is shown within 95% confidence intervals (blue region). The schematic of each experiment is shown in the upper left corner of each plot corresponding to the experiments in Figure 2.3.

2.3.3 Model Prediction

Our ultimate goal of building a quantitative model was to accurately predict the behavior of a newly designed experiment. To do this, we designed six new experiments that varied the basic elements of the 3-level cascade, and compared model-simulated trajectories to experimental SFGFP fluorescence time course trajectories of these experiments (Figure 2.5).

We first aimed to test the ability of the estimated parameters to capture changes in DNA template concentration of the cascade. Two experiments were designed that varied the concentration ratio of a single repressive connection of the cascade (Figure 2.5A), and the amount of antisense sRNA expressed from L3 in the full cascade (Figure 2.5B). Model predictions were made by running simulations using our estimated parameter sets with the parameters β_1 and β_2 multiplied by factors that took into account the change in DNA concentrations in the new experiments. In both new experiments, the average model predictions accurately matched the observed experimental trajectories, with all experimental trajectories falling within the 95% confidence intervals of the simulated trajectories (Figure 2.5A, B). In the single repression case of increasing L2 DNA, more repression was modeled and observed compared to the trajectories from the training experiment as expected (Figure 2.5A vs. Figure 2.4B). In the full cascade example, there were two sets of experimental trajectories observed, with the model showing predictions that captured the average behavior of these sets (Figure 2.5B).

Next we tested the ability to make predictions on RNA transcriptional cascades that included regulatory parts not involved in the training of any parameters. In particular, we designed an experiment to include a tandem attenuator in front of the L1 SFGFP reporter. Since the tandem attenuator is more sensitive to antisense RNA concentration, we expected both a single repressive connection and the full cascade to show an overall

lower SFGFP signal throughout the trajectory[108]. This was indeed observed (Figure 2.5C vs. Figure 2.4B and Figure 2.5D vs. Figure 2.4D). Furthermore, we found that the model predictions for these new circuit variants were in strong agreement with the experiments, with the average predictions accurately capturing the observed trajectories, for which all but one fell within the 95% confidence intervals of the simulated trajectories (Figure 2.5C,D). It is important to note that the tandem attenuator regulatory part was not included at all in the parameterization experiments. Rather it was modeled by squaring the repressive function of the single-attenuator (Figures 2.2C, A.1C, A.2B). This shows how our model is easily extensible, and that accurate predictions can be made from a limited set of parameterization experiments when modular parts are used to construct synthetic circuits.

Finally we tested the ability of our model to predict rewiring of the cascade elements. We designed two experiments that swapped the order of the R_1/A_1 and R_2/A_2 repressive sRNA/target attenuator pairs in both a single repressive connection and the full cascade (Figure 2.5E,F). This was a challenging prediction since the genetic contexts of R_1 and R_2 changed in the swapped configuration between tandem ribozyme-antisense constructs and single antisense expression. As we observed in the parameterization experiments, this context change causes large changes in the repression parameters for these antisense sRNAs (Table A.4). Therefore, we needed a way to easily estimate these new parameters using our previous information and as few as possible new parameterization experiments. To do this, we made several assumptions: (i) if the repression ratio of single copies of the two repressors, K_{1s} and K_{2s}, is f ($f = K_{1s}/K_{2s}$), then the repression ratio of tandem copies of the two repressors, K_{1d} and K_{2d}, is also f ($f = K_{1d}/K_{2d}$); (ii) the crosstalk strength of the swapped construct is the same as the crosstalk strength of the parameterization construct; and (iii) two different single repressors would have similar degradation and maturation rates, so that in the swap construct these two sets

of maturation and degradations rates are unchanged. These assumptions allowed us to re-use almost all parameters except K_{1s} and K_{2d}. Since K_{2s} and K_{1d} were already estimated in the parameterization procedure, we used one additional simple experiment to estimate K_{1s} (Figure A.5). This allowed us to calculate f, which could then be used to calculate K_{2d} by our assumptions. Finally K_{1s} and K_{2d} were used to make the prediction of our new design with the rest of the previously determined parameters. Comparisons between the average simulated trajectories were again in strong agreement with the observed SFGFP fluorescence trajectories for these experiments, with all trajectories falling within the 95% confidence intervals of the experiments (Figure 2.5E,F).

Overall we found that our model and its estimated parameters are capable of performing quantitative predictions of new sRNA circuits. This demonstrated its potential for aiding in the design of circuits for synthetic biology. In addition, we showed that our model is extensible, and can incorporate new genetic parts with limited additional characterization experiments.

2.3.4 Understanding TX-TL Batch Variation Using Model Parameterization

Given the simplicity and convenience of the four parameterization experiments, we wanted to investigate if these experiments could be used as a way to study batch-to-batch variation in TX-TL extract performance. In previous work, we observed differences between sRNA circuitry characterization time courses when different batches of TX-TL were used[108]. This was hypothesized to be due to different concentrations of molecular machinery in the extract that could impact the overall transcription, translation and degradation rates that influence circuit expression. Since our parameter esti-

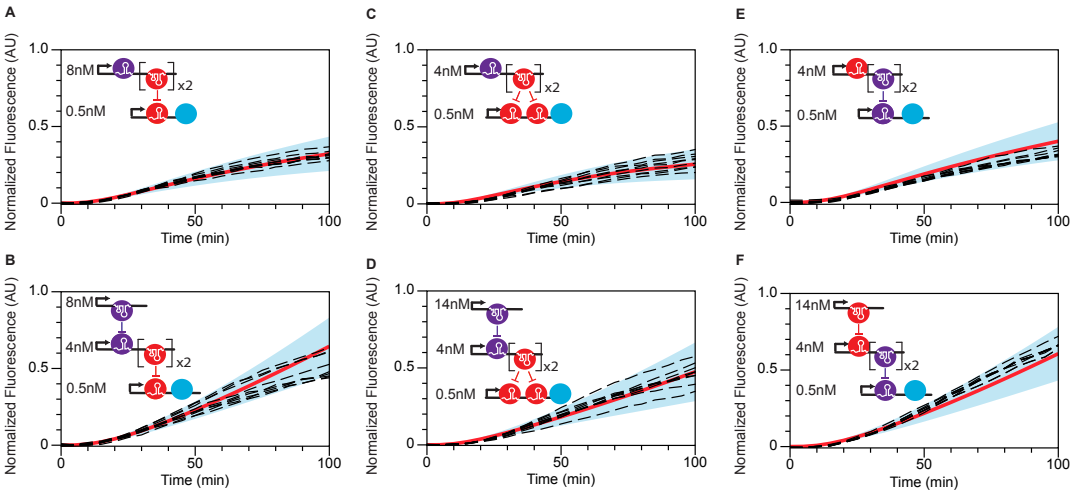


Figure 2.5: Comparison of experimental trajectories of SFGFP fluorescence in TX-TL experiments (black dash lines) with simulated model predictions. Model simulated trajectories were generated by performing 1,000 simulations with parameters drawn from the set of 10,000 determined from the estimation procedure (see Methods). Experimental and model trajectories were normalized by the maximum observed experimental fluorescence (see Methods). The mean simulated trajectory (red line) is shown within 95% confidence intervals (blue region). The schematic of each experiment is shown in the upper left corner of each plot. (A) Two level concentration prediction that varies the L2 plasmid concentration from Figure 2.3B. (B) Three level concentration prediction that varies the L3 plasmid concentration from Figure 2.3D. (C) Two level tandem attenuator prediction. The experiment contains 0.5nM of a modified L1 plasmid expressing 2 tandem copies of A_1 in front of SFGFP (L1T, Table A.2), and 4nM of the L2 plasmid. (D) Three level tandem attenuator prediction containing the same constructs as in (C) with an additional 14nM of L3 plasmid. (E) Two level swap prediction. The experiment contains 0.5 nM of a modified L1 plasmid that expresses A_2 in front of SFGFP (PL1, Table A.2), and 8nM of a new L2 plasmid expressing A_1 followed by R_2 (PL2, Table A.2). (E) Three level swap prediction containing the same constructs as in (D) with an additional 14nM of a new L3 plasmid that expresses R_1 (PL3, Table A.2).

mation procedure establishes quantitative values for each of these key rates, we thought that comparing estimated parameters from two different batches of TX-TL would more precisely reveal the specific differences between the batches.

To test this idea, we performed the same parameter estimation procedure with a separate TX-TL batch. This yielded the same estimates of all thirteen parameters in our model, which were shown to accurately model the new experimental SFGFP circuit

characterization trajectories (Figure A.6). Although these two batches showed comparable quality, we noticed that batch A had slightly lower GFP fluorescence trajectories compared to batch B in three out of four parameterization experiments (Figure A.7). To analyze this batch-to-batch difference, we compared the distributions of the thirteen parameters derived from these two sets of experiments (Figure 2.6, Table A.4).

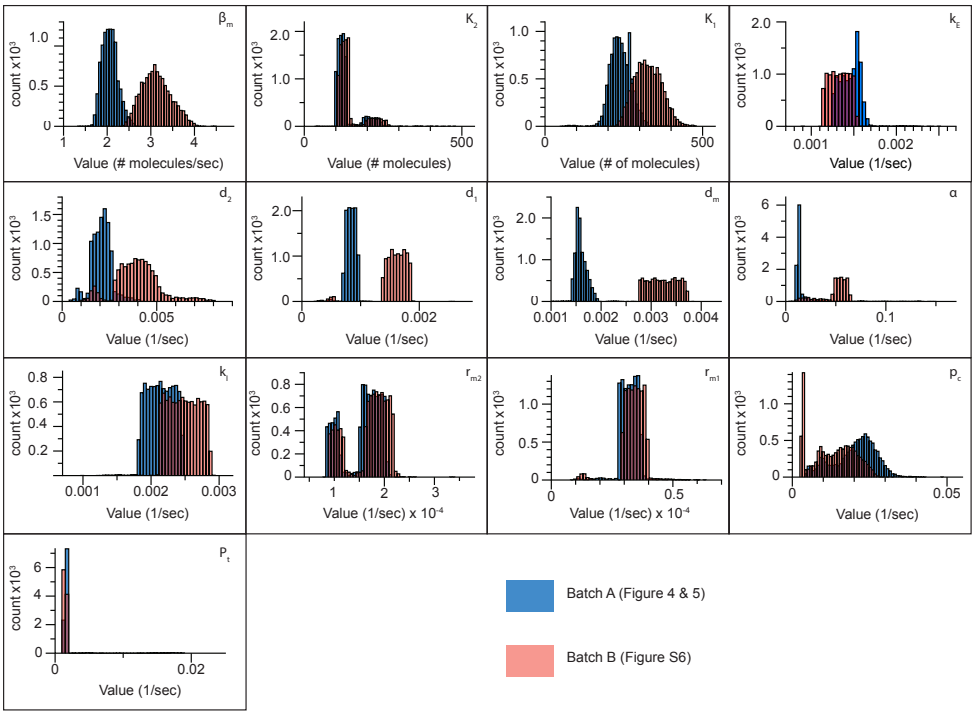


Figure 2.6: Parameter distributions from two independent TX-TL batches. Estimated parameters from batch A (blue) (Figure 2.4) and batch B (red) (Figure A6). Histograms are composed of 10,000 sets of parameters fit from the parameterization experiments performed in each batch.

Interestingly, we found that several parameters showed large differences between the two batches. These included the transcription rate β_m, RNA degradation rates (d_2, d_1, d_m) and SFGFP maturation rate . Specifically, batch A had lower values for all of these parameters, suggesting that batch B had faster transcription, RNA degradation and GFP maturation. This could be explained by key differences in molecular machinery

concentrations between these two batches that are known sources of variability in extract preparation[106]. Furthermore, these results demonstrated that our rapid parameter estimation procedure can be used to generate hypotheses about specific differences between TX-TL extract preparations, which could become increasingly important given the emerging new applications for these systems as molecular diagnostic platforms[84].

2.4 Conclusion

In conclusion, we have developed an effective quantitative model for sRNA transcriptional circuits and demonstrated its accuracy using a 3-level sRNA repression cascade as a test case. Our results showed that we were able to capture key features of this mechanistically sophisticated circuitry using only 8 ODEs with thirteen unknown parameters. To determine these parameters, we used sensitivity analysis to design four simple experiments that can be performed in parallel using cell-free TX-TL extracts to estimate all thirteen parameters. Finally we used our model along with estimated parameters to predict the time course dynamic trajectories for new network designs that used parts that were not included in parameterization experiments. We also showed that our model was easily extensible to include new parts with a limited number of additional characterization experiments. In all cases, our models were able to accurately reproduce the experimental results.

Interestingly, the process of constructing, parameterizing and validating our model of sRNA transcriptional circuitry revealed several new features of TX-TL systems and the sRNA regulators used. In particular, we found that there was a need to introduce a TX-TL pre-incubation step in order for computational analysis to match observed experimental trajectories. This is presumably due to certain processes that need to occur for the extract system to become active. In addition, we also discovered that the sRNAs used in our circuitry do not appear to be immediately active once synthesized, and we had to include equations that captured a maturation process to accurately model experimental SFGFP trajectories.

Several aspects of this work are significant. First, this represents the first validated computational model of synthetic RNA transcriptional genetic circuitry. Combined with

other models that capture the effects of sRNA translational regulation[50], this work helps lay the foundation for CAD tools that can incorporate RNA regulators into synthetic circuitry design[26]. We anticipate this to be more important as researchers increasingly turn to RNA-mediated gene regulatory systems for controlling gene expression in biological systems.

Second, the sensitivity analysis-based parameterization procedure is completely general, and could find wide use for establishing parameters for many synthetic regulators and circuits. In particular, the same exact procedure could be used to find parameters for the wide array of synthetic regulators now at our disposal[76], including the exciting CRISPRi-based regulatory mechanisms. While we used the speed and experimental convenience of TX-TL reactions to perform our circuit parameterization experiments, this method should be extensible to models that capture the behavior of networks *in vivo*. In addition, this methodology could become even more powerful when coupled to different methods for implementing parameterization experiment design. For example, while we considered experimental design in terms of including different concentrations of DNA templates in the TX-TL reactions, one could easily imagine applying optogenetic control of component activity to design complex experiments that could allow even more efficient parameter estimation[83]. In these systems, different patterns of input light could be used to drive the system in specific ways designed to obtain the maximum number of parameters in the minimum amount of time.

Finally, we showed that our overall modeling and parameterization procedure offers a new approach for studying the underlying causes of batch-to-batch variation in TX-TL systems. In particular we found that different batches of TX-TL led to different distributions for key parameters that are directly related to the concentrations of core transcription/translation machinery in the batches (Figure 2.6). Understanding the basis

of TX-TL variation will become increasingly important as these systems find wide use for a variety of applications including as metabolic production systems[44], for rapid prototyping and characterization of genetic circuits[106, 108], and for new types of molecular diagnostics[84].

2.5 Methods

2.5.1 Sensitivity Analysis to Design Parameterization Experiments

Our model for RNA circuitry consists of a set of ordinary differential equations that describe the time varying rate of change in the concentrations of the molecular species that participate in the circuitry, $x_i(t)$ (Figure 2.2). These equations are parameterized by a set of parameters, p_j, that we want to estimate by fitting model predictions to a small set of experiments. These experiments were designed through an iterative process of sensitivity analysis on the set of model equations (Figure 2.2).

An individual experiment was considered to be a TX-TL reaction containing a subset of the DNA constructs encoding the full three-level sRNA transcriptional cascade at defined concentrations. Each such experiment produces a measureable trajectory of SFGFP fluorescence as a function of time, and can be modeled by the subset of equations that describe the gene expression processes from the included DNA. For example, if the TX-TL reaction contains only L1, then only the last four equations in Figure 2.2 need to be used with R_1=R_2=0. After a specific experiment (subset of DNA) was proposed, the next step was to assess which parameters were identifiable from this experiment, which is closely linked with parametric sensitivity analysis. Here we used the procedure proposed by McAuley and coworkers[127] to first calculate and analyze the sensitivity coefficient matrix for the proposed experiment as follows.

For each experiment, the sensitivity coefficient matrix $z_{ij}(t)$, is a time-varying matrix that encapsulates how sensitive the concentration of the molecular species x_i is to a change in the parameter p_j

$$Z_{ij}(t) = \left.\frac{\partial x_i}{\partial P_j}\right|_t \qquad i = 1, 2, ...N \qquad j = 1, 2, ..., P$$

Here P denotes the number of parameters and N denotes the number of molecular species. If we write the model equations generally as

$$\frac{dx_i}{dt} = f_i(\mathbf{x}, \mathbf{P}, t)$$

then it can be shown that $Z_{ij}(t)$ are the solutions to a set of differential equations given by

$$\frac{dz_{ij}}{dt} = \sum_{N}^{k=1} \frac{\partial f_i}{\partial x_k}(\mathbf{x}, \mathbf{P}, t) z_{k,j} + \frac{\partial f_i}{\partial p_j}(\mathbf{x}, \mathbf{P}, t)$$

which are subject to the initial condition $z_{ij} = 0$. Since our only observable in the TX-TL experiment is SFGFP, we focused specifically on $z_{SFGFP,j}(t)$ to determine which parameters were identifiable in the experiment.

To begin the experimental design process, we first determined a set of parameters that closely matched experiments by hand-fitting the parameters against SFGFP trajectories measured from TX-TL reactions containing subsets of the cascade DNA elements, using initial guesses based on the literature findings[116, 129, 126, 86]. We next proposed the simplest experimental design (a fixed concentration of the bottom level L1) to the sensitivity analysis procedure. Using the hand-fit parameters, $z_{SFGFP,j}(t_k)$ was calculated by solving the equations shown in Figure 2.2 using Matlab over a set of discrete time steps, t_k, and then scaled by multiplying by $p_j / x_{SFGFP}(t_k)$. Identifiability was then performed according to McAuley[127].This was done by finding the column of this matrix that had the biggest magnitude (indicating the most sensitive parameter), calculating a residual matrix which removed this column and controlled for correlations between parameters, and iterating this procedure on the resulting residual matrix until a threshold was reached on the largest remaining column magnitude. In this way a set of parameters was determined that maximally influenced the modeled trajectory of the proposed experiment (Figure 2.3A).

After performing this procedure on the simplest experiment (L1), we proposed a fur-

ther experiment and performed the same analysis, except that parameters already identified by previous experiments were marked as determined by setting their columns in the sensitivity matrix to 0. Rounds of experimental design and sensitivity analysis were performed until all 15 parameters were able to be identified by four TX-TL experiments (Figure 2.3).

2.5.2 Parameter Estimation

Parameters were estimated from each designed experiment by fitting the identifiable parameters of that experiment to measured SFGFP expression time trajectories. The parameter estimation problem is given by:

$$min(G_{t_k}^{experimental,mean} - G_{t_k}^{simulated})^2$$

Where $G_{t_k}^{experimental,mean}$ denotes the average value of the experimentally observed SFGFP expression at a certain time t_k. The vector P contains all of the identifiable parameters in the experiment being analyzed. $G_{t_k}^{simulated}(P)$ denotes the model simulated SFGFP expression at time t_k. For a specific experiment, optimal P vectors were found using the *Matlab* function *fmincon*. Certain parameters were constrained to lie within specific values. For example, the repression concentrations K_1 and K_2 were constrained to be close in magnitude to each other based on the known similarity in repression of the two sRNA transcription repressor variants[64]. A complete list of constraints used is in Table A.3.

Sets of identifiable parameters were estimated from the corresponding experimental trajectories in turn. We first used the estimation procedure to optimize the initial guesses used in the identifiability analysis above. To do this, we constructed uniform distributions (100 points) around each parameter value (±15%), generating 100 sets of

parameters. Each set of parameters served as a different starting point for finding optimal parameter estimates. Each estimate set was found by sequentially applying the fitting procedure above to each of the designed experiments, only fitting the identifiable parameters for that experiment. Fit parameters from one experiment were then used to update and replace the initial guesses before moving on to the next experiment until each of the 13 parameters was fit from the four experiments. This was repeated for each initial set of parameters to produce 100 sets of estimated parameters. We then chose the set that produced the closest simulated trajectory compared to the experimental data and used this as the guessed parameter set for the next iteration. We repeated this process 10 times. The final optimal parameter set was used in the same procedure to generate 10,000 sets of estimated parameters, which were then subject to the analysis outlined in the main text.

To plot our results, we scaled $G_{t_k}^{simulated}$ by the observed experimental values according to:

$$\hat{G_{t_k}}^{simulated} = \frac{G_{t_k}^{simulated} - G_{t_0}^{simulated}}{G_{max}^{experiment} - G_{t_0}^{experiment}}$$

where $G_{t_k}^{simulated}$ and $G_{t_0}^{simulated}$ are the SFGFP expression level at any time t_k and initial time t_0,respectively,$G_{max}^{experiment}$ is the maximum experimental data point of the entire trajectories and $G_{t_0}^{experiment}$ is the experimental data point at t=0. $G_{t_k}^{experiment}$ was also scaled as:

$$\hat{G_{t_k}}^{experiment} = \frac{G_{t_k}^{experiment}}{G_{max}^{experiment} - G_{t_0}^{experiment}}$$

where $G_{t_k}^{experiment}$ is the experimental expression of SFGFP at any time t_k. In order to make all trajectories comparable, all experimental and modeled trajectories were scaled by the experimentally observed SFGFP fluorescence value of the first parameterization experiment at 100 minutes (Figure 2.4A).

2.5.3 Plasmid construction and purification

A table of all the plasmids used in this study can be found in Supporting Table A.2, with key sequences found in Supporting Table A.1. The pT181 attenuator and repressor plasmids, pT181 mutant attenuator and antisense plasmids, and the no-antisense control plasmid were constructs pAPA1272, pAPA1256, pAPA1273, pAPA1257, and pAPA1260, respectively, from Lucks et al.[64]. The second level of the cascade (JBL069) was modified from construct pAPA1347 from Lucks et al.[64]. The double attenuator and modified level 2 constructs for prediction were created using Golden Gate assembly[32]. Plasmids were purified using a Qiagen QIAfilter Plasmid Midi Kit (Catalog number: 12243) followed by isopropanol precipitation and eluted with double distilled water.

2.5.4 TX-TL extract and buffer preparation

Extract preparation:

Cell extract and reaction buffer was prepared according to Shin and Noireaux[98] and Sun et al.[106]. In brief, *E.coli* BL21 Rosetta cells were grown to an OD600 of 1.5, pelleted via centrifugation, and washed with a buffer at pH 7.7 containing Mg-glutamate, K-glutamate, Tris and DTT. Lysis was performed via bead-beating, followed by centrifugation to remove beads and cell debris. The resulting supernatant was incubated at 37°C for 80 minutes and then centrifuged, to remove endogenous nucleic acids. The supernatant was dialyzed against a buffer at pH 8.2, containing Mg-glutamate, K-glutamate, Tris and DTT. The extract was then centrifuged, and the supernatant flash-frozen in liquid nitrogen and stored at -80°C. The cell extract for Batch A had a protein concentration

of 30 mg/mL, and its expression was optimized via the addition of 1 mM Mg and 40 mM K. For Batch B: 30 mg/mL protein,1 mM Mg, and 80 mM K.

Buffer preparation:

The reaction buffer was prepared according to Sun et al.[106], and consists of an energy solution (HEPES pH 8 700 mM, ATP 21 mM, GTP 21 mM, CTP 12.6 mM, UTP 12.6 mM, tRNA 2.8 mg/ml, CoA 3.64 mM, NAD 4.62 mM, cAMP 10.5 mM, Folinic Acid 0.95 mM, Spermidine 14 mM, and 3-PGA 420 mM) and amino acids (leucine, 5 mM, all other amino acids, 6 mM). Extract and buffer were aliquoted in separate tubes (volume appropriate for seven reactions) and stored at -80°C.

2.5.5 TX-TL Experiment

TX-TL buffer and extract tubes were thawed on ice for approximately 20 min. Separate reaction tubes were prepared with combinations of DNA representing a given circuit condition. Appropriate volumes of DNA, buffer, and extract were calculated using a custom spreadsheet developed by Sun et al.[106]. Buffer and extract were mixed together and incubated for another 20 min at 37°C. DNA for the specific experiment was then added into each tube according to the previously published protocol[106]. Ten μL of each TX-TL reaction mixture was transferred to a 384-well plate (Nunc 142761), covered with a plate seal (Nunc 232701), and placed on a Biotek SynergyH1m plate reader. We note that special care was needed when pipetting to avoid air bubbles, which can interfere with fluorescence measurements. Temperature was controlled at 29°C. SFGFP fluorescence was measured (485 nm excitation, 520 emission) every five min for 100 min. Each reaction was run with a minimum of triplicate repeats, and repeated three

times independently (minimum of nine total replicates). A ten L sample of each TX-TL buffer and extract mixture was run together with each independent reaction as blank. All data points were then processed with blank values subtracted.

CHAPTER 3

ENGINEERING A FUNCTIONAL SMALL RNA NEGATIVE AUTOREGULATION NETWORK WITH MODEL-GUIDED DESIGN [1]

3.1 Abstract

RNA regulators are powerful components of the synthetic biology toolbox. Here, we expand the repertoire of synthetic gene networks built from these regulators by constructing a transcriptional negative autoregulation (NAR) network out of small RNAs (sRNAs). NAR network motifs are core motifs of natural genetic networks, and are known for reducing network response time and steady state signal noise. Here we use cell-free transcription-translation (TX-TL) reactions and a computational model to design and prototype sRNA NAR constructs. Using parameter sensitivity analysis, we design a simple set of experiments that allow us to accurately predict NAR function in TX-TL. We transfer successful network designs into *Escherichia coli* and show that our sRNA transcriptional network reduces both network response time and noise in steady-state gene expression. This work broadens our ability to construct increasingly sophisticated RNA genetic networks with predictable function.

3.2 Introduction

A major goal of synthetic biology is to design and construct synthetic gene networks that can be used to reprogram the behavior of living cells to accomplish specialized

[1] THIS WORK WAS ORIGINALLY PUBLISHED IN ACS SYNTHETIC BIOLOGY AND HAS BEEN REPRODUCED HERE WITH PERMISSION. HU, C. Y., TAKAHASHI, M. K., ZHANG, Y., & LUCKS, J. B. (2018). ENGINEERING A FUNCTIONAL SMALL RNA NEGATIVE AUTOREGULATION NETWORK WITH MODEL-GUIDED DESIGN. ACS SYNTHETIC BIOLOGY, 7(6), 15071518. HTTP://DOI.ORG/10.1021/ACSSYNBIO.7B00440

functions. RNA regulators have emerged as powerful tools in the synthetic biology toolbox of gene expression control that can be used to achieve this goal[20]. Specifically, synthetic biologists have engineered a diverse array of RNA regulators that can be used to control many of the core processes of gene expression including transcription[64, 20, 89, 22], translation[54, 40, 85] and mRNA degradation[17, 16, 33, 121]. In addition to being able to tune the expression of individual genes, these engineered RNA regulators have also been connected together to create synthetic gene networks in the form of logic gates [64, 20, 40, 119, 22], cascades[64, 78, 60, 108, 119, 22, 124], single input modules[108, 22] feed forward loops[22], and hybrid feedback controllers[121]. While these synthetic network motifs showcase some of the advantages of RNA regulators, they represent the beginning of larger efforts to create a greater diversity of RNA network function that is necessary to engineer higher order cellular behavior. Much of the design of new synthetic genetic networks has been guided by the understanding and repurposing of natural genetic networks that control a large range of cellular functions including cellular maintenance, development, and the response to environmental perturbations. These natural genetic networks are composed of repeating occurrences of simple patterns called network motifs[1]. Examples of network motifs include autoregulation where regulators can regulate their own synthesis, feedback loops where regulators can regulate upstream portions of a network, and feed-forward loops where regulators can regulate downstream portions of a network, among others[1]. Each of these network motifs has a distinct information processing function, and larger networks are made by composing motifs in different ways to achieve higher order information processing. In essence, these network motifs form a basis set of elements from which larger networks can be built, making the construction and characterization of synthetic versions a powerful target for synthetic biology[1].

The simplest motif in transcription networks is negative autoregulation (NAR)[1],

a motif in which transcription factors negatively regulate their own transcription[93]. One of the key functions of NAR networks in nature is to speed the response time of a transcription network[91]. NAR networks are also widely believed to have noise reduction properties[7, 66, 75, 97, 99]. At the same time, NAR networks are a key element of genetic oscillator modules[80] as well as biological feedback controllers[29]. Thus NAR networks represent a rich target to construct using RNA genetic circuitry, which may have particular advantages due to the fast dynamics observed for RNA mediated networks[109].

In this work, we engineer the first sRNA transcriptional NAR network using the transcriptional attenuator from the Staphylococcus aureus plasmid pT181[81]. To do this, we leverage recent advances in using cell-free transcription-translation (TX-TL) reactions[108] and a computational modeling framework[48] to rapidly prototype and characterize NAR constructs that are then transferred to living cells. We first simulate RNA-based NAR networks qualitatively with a simplified model to anticipate challenges of building an RNA-based NAR network. Following the models suggestion that increased copies of the RNA repressor would be needed for observable NAR function, we construct and optimize new RNA parts to construct the NAR network. We then develop a quantitative model of the network by representing the biological processes as a system of ordinary differential equations (ODEs), and use a sensitivity analysis approach to design experiments to fit the parameters of the model. Next, we estimate all unknown parameters of the NAR network and predict its dynamic trajectories in TX-TL reactions quantitatively. Comparison of the model with experimental data shows that the model captures the experimental results within a 95% confidence interval. Moreover, we show that the NAR network speeds up the response time for both modeled and experimental dynamic trajectories. Finally, we move the TX-TL-prototyped NAR network into *E. coli*, where we show that our NAR network also significantly reduces network response

time as well as steady state signal *in vivo*.

3.3 Results and Discussion

3.3.1 Using simplified computational models to guide NAR network design

To build an RNA-based NAR network, we sought to use the pT181 transcriptional attenuator[81] which we have previously used to build RNA-only transcriptional networks[64, 109, 20]. The pT181 attenuator is an RNA sequence in the 5′ untranslated region of a gene that regulates transcription elongation through the formation of an intrinsic terminator hairpin (Figure 3.1A). In its native or ON state, intramolecular interactions with an anti-terminator sequence prevents terminator formation and allows transcription elongation. In its OFF state, interactions with a repressor sRNA, which is complementary to the upstream region of the attenuator, sequesters the anti-terminator sequence, allowing the terminator hairpin to form and prevent transcription elongation of the downstream gene (Figure B.1A)[81, 12]. Thus, the combination of repressor sRNA and attenuator act as an RNA-based transcriptional repressor.

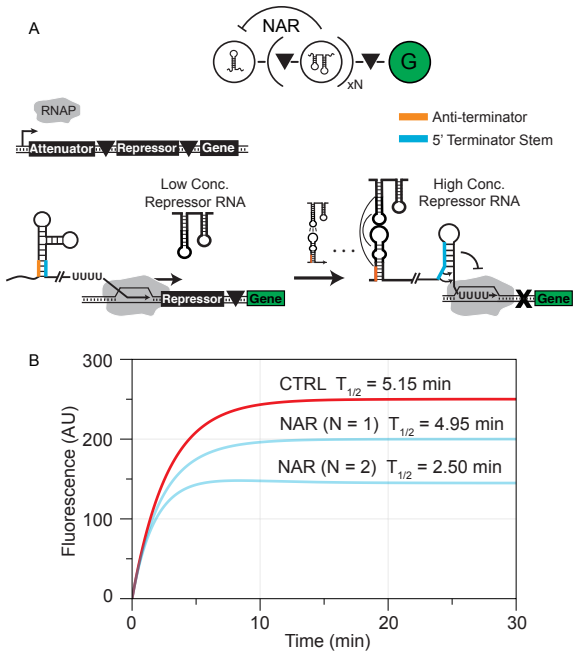
To build an NAR network, we first proposed a construct in which the pT181 attenuator regulates the transcription of its own repressor sRNA, followed by a reporter gene (Figure 3.1A). At the start, the attenuator would allow transcription of both the repressor sRNA and reporter gene until the concentration of repressor sRNA builds up and binds to the attenuator. Interaction of the repressor sRNA with the attenuator would lead to terminator formation and reduce further transcription of repressor sRNA and the reporter gene.

We first sought to determine the feasibility of this pT181 attenuator based NAR network design. To do this, we developed a simple mathematical model as an empirical

guide for building this network. The model used parameters from previous work where we simulated a transcriptional cascade in TX-TL using the pT181 attenuator[48]. Since the RNA-based NAR would uniquely operate at the RNA level, we sought a reporter that would allow us to directly measure and model RNA levels within the system. We therefore proposed an NAR construct that used the malachite green (MG)-binding RNA aptamer as a reporter[39]. Once folded properly, the RNA aptamer can bind the small molecule chromophore malachite green (MG), which results in a fluorescence output. The MG aptamer, therefore, allows experimental characterization of the RNA levels of the network with a convenient fluorescence read out[39], and would allow us to directly model time trajectories of our experimental observable.

Our initial qualitative model treated the basic gene expression processes of the NAR network as coupled chemical reactions of two species: the repressor sRNA (R) and the reporter malachite green RNA (Figure 3.1B). We considered the concentrations of these two species in time to be functions of their synthesis and degradation rates. For simplicity, we modeled the transcriptional repression with a first order Hill function based on previous experimental work[64], in which K represents the concentration of active repressor that is required to achieve significant repression of downstream gene transcription (Figure 3.1B). Finally, to simulate the model, we used estimated parameter values based on parameters we obtained in previous work in modeling RNA-based transcriptional cascades in TX-TL reactions[48] (Table B.1).

One of the defining characteristics of an NAR network is that it speeds up a networks response time, defined as the time it takes for a dynamic process to reach half of its steady state level[91]. We therefore used response time as an indicator of a functional NAR network. The initial simulation of the NAR network showed an insignificant decrease in response time compared to a control (4.95 min compared to 5.15 min) (Fig-



$$\frac{dR}{dt} = \beta \cdot \left(1 - \frac{R}{K+R}\right) - d \cdot R$$

$$\frac{dG}{dt} = \beta \cdot \left(1 - \frac{R}{K+R}\right) - d_m \cdot G$$

Figure 3.1: A simple model of the proposed sRNA transcriptional NAR construct uncovers potential challenges and provides design guidelines for network construction. (A) The transcriptional NAR network motif and proposed sRNA implementation. The pT181 attenuator regulates the transcription of N copies of its own repressor RNA (repressor, R) and a reporter gene (G), each insulated from each other by self-cleaving ribozymes (triangles). Once transcription is activated, initially RNA polymerase is allowed to transcribe the repressor RNA-reporter gene construct. As concentration of repressor RNA increases, it binds to the attenuator, which leads to intrinsic terminator formation and repression of transcription of the repressor RNA-reporter gene construct. (B) A simplified two-component model that describes the synthesis and degradation of the repressor (R) and reporter gene (G) provides insight for network construction. The dynamic trajectories were simulated by this model with parameters adapted from a previous study (Table B.1)[48]. The simulation suggests that a single copy of repressor RNA ($N = 1$) would be insufficient to significantly reduce network response time, while tandem copies of repressor RNA ($N = 2$) would.

ure 3.1B). We hypothesized this was due to the fast degradation rate of repressor sRNA and weak repression of the attenuator that prohibits sufficient accumulation of repressor sRNA to achieve significant repression. To solve this potential problem, we proposed a new NAR design that increased the concentration of repressor sRNA by including two copies of the RNA in tandem. This was motivated by previous work that showed that two repressor sRNAs in tandem result in stronger repression [64, 108, 48] (Figure B.2A), and longer sRNA transcripts (two repressors) degrade slower than a shorter (one repressor) version [48] (Figure B.2B). Therefore, for convenience, in the simple model we treated the two-copy configuration as a new repressor sRNA species that is twice as repressive, but that also has a degradation rate which is half of that of a single repressor RNA to account for its increase in size. The simulation of this construct suggested that by using two tandem copies of repressor sRNA, the response time would decrease from $T_{1/2} = 5.15$ minutes to $T_{1/2} = 2.50$ minutes, which would also result in an observable decrease in fluorescent signal (Figure 3.1B).

Our qualitative model thus allowed us to capture the essence of an RNA-based NAR network and predict qualitative response time changes for different network configurations. Based on the guidance of this simplified model, we next sought to implement an RNA-based NAR network with tandem copies of repressor sRNAs.

3.3.2 Optimization of RNA parts for NAR network construction

Overall, the proposed NAR architecture consists of four distinct RNA parts that must be fused together in tandem on the same RNA molecule for proper function: an upstream attenuator to control transcription, two copies of the repressor sRNA to provide the negative feedback if synthesized, and the MG aptamer to provide a fluorescent read-out

(Figure 3.4A). In addition, there are accessory transcriptional terminators on the ends of constructs that are needed to define expression constructs. A functional NAR requires each individual RNA part to fold into proper structures, which is increasingly difficult with more RNA parts in tandem. To help alleviate potential problems with misfolding and mis-function, we therefore sought to minimize the size of the parts involved in the construct. Specifically, we sought to minimize the length of the repressor sRNA required for function and the length of accessory terminators at the end of constructs to remove unnecessary sequence.

Since the accessory terminators are the largest individual part in the proposed NAR network, we first sought to replace the rrnB terminator (TrrnB) originally used in the development of the pT181 attenuator synthetic construct[64] with a structurally smaller terminator, and then to reduce the repressor sRNA to the minimal structure necessary for function. The rrnB terminator consists of 368 nucleotides and is predicted to fold into a complex structure that could be a source of unwanted RNA-RNA interactions in our NAR network. In previous work, we have shown that the rrnB terminator did in fact interfere with the function of a loop-linear chimeric attenuator[110]. Recently, *Chen et al.*[24] characterized 582 natural and synthetic transcriptional terminators, which we used as a source for possible terminator alternatives. We chose five terminators that had compact predicted secondary structures and similar or better termination efficiency than that of the rrnB terminator. The terminators were first tested in a simple context to ensure the repressive strength and orthogonality of the repressor RNA were not affected by changing the terminator. To test this, we placed each terminator downstream of the wild type pT181 repressor sRNA, which was constitutively expressed on a high copy plasmid. Each repressor sRNA-terminator construct was tested for repression of the wild type attenuator and its orthogonality to a mutant attenuator *in vivo*, which were placed downstream of the same constitutive promoter and upstream of the super-folder green

fluorescent protein (SFGFP) coding sequence[86] on a medium copy plasmid (Figure S1). Attenuator and repressor sRNA pairs were transformed into *E. coli* TG1 cells, and fluorescence (485 nm excitation, 520 nm emission) and optical density (OD) (600 nm) were measured for each culture (see Methods). Of the five terminators tested, the repressor sRNA-L3S3P21 terminator construct had the best combination of repression of the wild type attenuator and orthogonality to the mutant attenuator (Figure B.1) and was therefore chosen for use in the NAR network.

Next, we tested minimizing the repressor sRNA construct. Early *in vitro* work with the pT181 attenuator showed that the full-length wild type repressor RNA was not required for interaction with and repression of the attenuator[11]. We therefore tested a truncated version of the repressor sRNA, which only contained the hairpin thought to be involved in the initial kissing hairpin interaction with the attenuator. The single hairpin repressor RNA followed by the L3S3P21 terminator showed equivalent repression (90%) of the attenuator when compared to the full-length repressor RNA (Figure 3.2A).

In the final NAR construct, it is essential that the individual repressor sRNA copies are free to fold into functional conformations and repress the attenuator. It was previously shown that using the hammerhead ribozyme from small tobacco ring spot virus (sTRSV)[59] to separate tandem copies of repressor RNA on the same transcript improved function of the repressor RNA constructs[64]. Since only the full-length repressor RNA had been tested with the sTRSV ribozyme, we sought to confirm that the sTRSV ribozyme could be used to separate copies of the single hairpin repressor RNA without affecting function. We tested constructs with the sTRSV ribozyme both upstream and downstream of the single hairpin repressor sRNA as well as a tandem double repressor sRNA construct and showed that the sTRSV ribozyme does not interfere

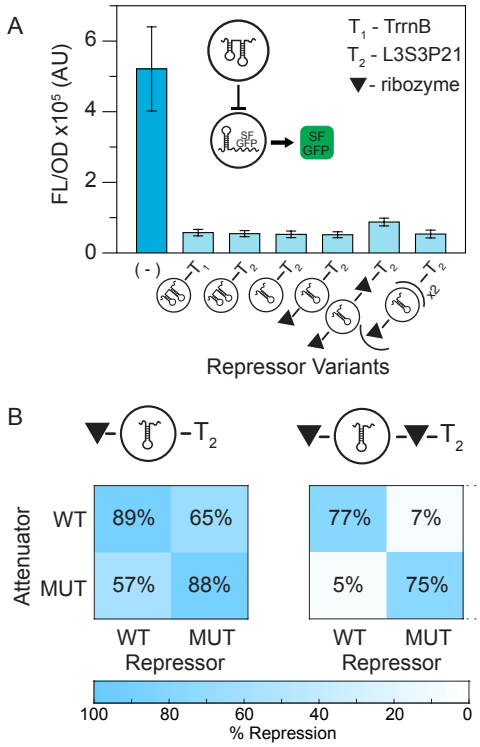


Figure 3.2: Optimization of RNA parts for network construction. (A) Average steady-state fluorescence (FL/OD) of *E. coli* TG1 cells measuring repression of a construct consisting of the pT181 attenuator controlling SFGFP expression. The pT181-SFGFP construct was transformed with a control plasmid (-), or constructs that expressed variations of the repressor RNA. Variations in the terminator included downstream of the repressor tested the impact of the size of the terminator (T1 vs. T2), with L3S3P2134 (T2) being smaller than TrrnB (T1). Variations in the size of the repressor RNA repressor (two hairpins vs. one hairpin) tested the impact of including reduced regions of the repressor RNA on repressive function17. Other variations tested the impact of using the sTRSV ribozyme (triangles) for separating RNA parts and including multiple copies of the repressor. Error bars represent standard deviations over nine biological replicates. (B) Orthogonality matrices illustrating that the single hairpin repressor RNA followed by terminator L3S3P21 achieves adequate repression but compromises orthogonality (left). Matrices were generated by measuring the fluorescence of cells transformed with different combinations of the wild type (WT) or specificity mutated (MUT) pT181 attenuator-SFGFP/repressor expression constructs in the repressor configurations indicated by the schematic. Adding a ribozyme in between the single hairpin repressor RNA and terminator L3S3P21 restores the orthogonality (right). Each box in the matrices represents % repression of cells with a repressor RNA expression plasmid compared to cells with the control plasmid. Repression is presented by a color scale in which 100% is blue and 0% is white. Repression % was calculated from the average over nine biological replicates.

with the single hairpin repressor sRNA function (Figure 3.2A).

Interestingly, while the sTRSV-single hairpin repressor sRNA-L3S3P21 terminator construct showed equivalent repression of the attenuator compared to the construct without the ribozyme, the ribozyme construct was no longer orthogonal to a mutant pT181 attenuator (Figure 3.2B). However, we observed that adding a second ribozyme downstream of the single hairpin repressor sRNA and before the terminator recovered orthogonality (Figure 3.2B). Overall, these optimizations allowed us to remove 387 nucleotides (36% of the total length) from the NAR design. We thus continued constructing and analyzing the sRNA transcriptional NAR with these minimized parts.

3.3.3 A quantitative model accurately predicts NAR network function in TX-TL

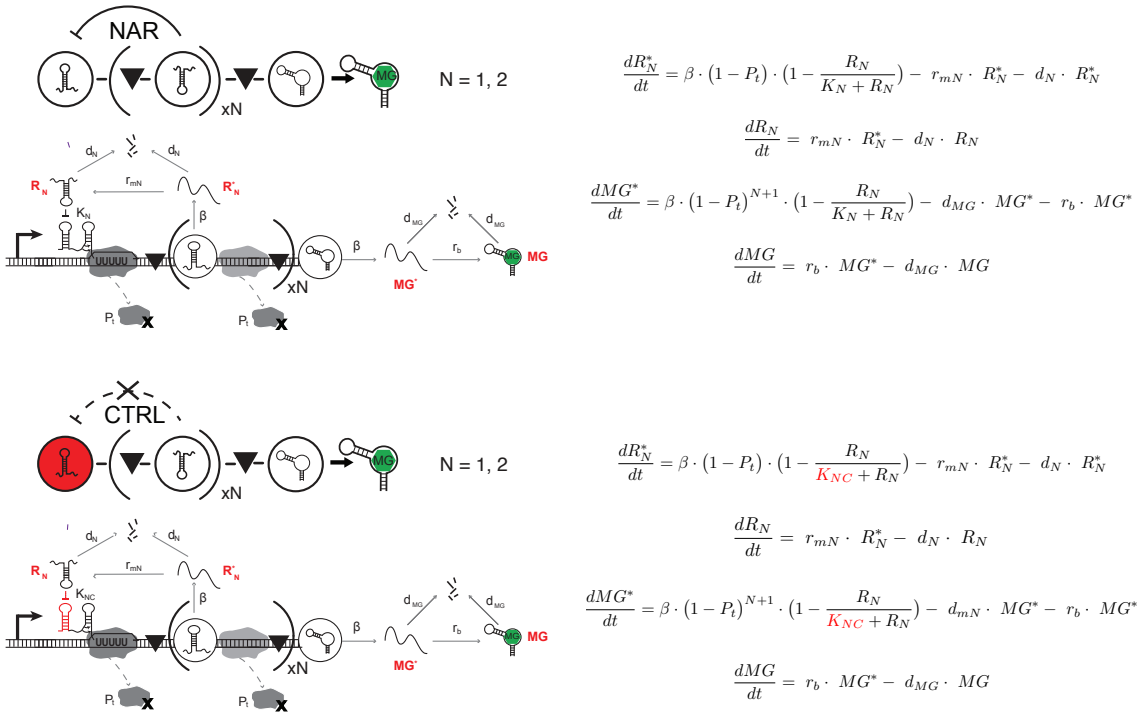
Our next goal was to build a more accurate model that could provide quantitative predictions of the NAR network function in TX-TL reactions. To do this, we expanded the set of ODE equations that modeled repressor R and reporter MG in Figure 1 to include the maturation steps for both R and MG (Figure 3.3). Specifically, our previous work showed that time trajectories of TX-TL reactions could only be accurately modeled for the pT181 system if the maturation of the repressor sRNA was included[48]. This was explicitly included in the model by considering a maturation step whereby the initially transcribed immature species (R^*) folds into its functional structure (R). Similarly, initially transcribed MG molecules (MG^*) would need a further step to bind the dye before becoming fully matured and observable (MG). Both of these maturation processes were modeled by introducing additional species and specific rates of maturation (Figure 3.3). For the MG aptamer, we modeled both folding and dye binding with one term, r_b, to

Table 3.1: NAR model parameters

Parameter	Description	Unit
K_1	Repression coefficient of single repressor.	Concentration (Conc.)
K_2	Repression coefficient of double repressor.	Conc.
K_{1C}	Repression coefficient of single mismatched repressor due to crosstalk.	Conc.
K_{2C}	Repression coefficient of double mismatched repressor due to crosstalk.	Conc.
d_1	Degradation rate of single repressor.	1/sec
r_{m1}	Maturation rate of single repressor.	1/sec
r_{m2}	Maturation rate of double repressor.	1/sec
r_b	Maturation and binding rate of malachite green (MG) aptamer.	1/sec
β	Transcription rate of pre-cleaved MG RNA.	Conc./sec
d_{MG}	Degradation rate of MG RNA.	1/sec
d_2	Degradation rate of double repressor.	1/sec
P_t	Probability of transcriptional disruption.	Unitless

Table 3.2: NAR model species

Model Species	Definition	Unit
R_1^*	Immature single sRNA repressor.	Conc.
R_1	Mature single sRNA repressor.	Conc.
R_2^*	Immature tandem sRNA repressor.	Conc.
R_2	Mature tandem repressor.	Conc.
MG^*	Unfolded/unbound MG aptamer.	Conc.
MG	Folded MG aptamer bound with dye.	Conc.



$$\frac{dR_N^*}{dt} = \beta \cdot (1 - P_t) \cdot \left(1 - \frac{R_N}{K_N + R_N}\right) - r_{mN} \cdot R_N^* - d_N \cdot R_N^*$$

$$\frac{dR_N}{dt} = r_{mN} \cdot R_N^* - d_N \cdot R_N$$

$$\frac{dMG^*}{dt} = \beta \cdot (1 - P_t)^{N+1} \cdot \left(1 - \frac{R_N}{K_N + R_N}\right) - d_{MG} \cdot MG^* - r_b \cdot MG^*$$

$$\frac{dMG}{dt} = r_b \cdot MG^* - d_{MG} \cdot MG$$

$$\frac{dR_N^*}{dt} = \beta \cdot (1 - P_t) \cdot \left(1 - \frac{R_N}{K_{NC} + R_N}\right) - r_{mN} \cdot R_N^* - d_N \cdot R_N^*$$

$$\frac{dR_N}{dt} = r_{mN} \cdot R_N^* - d_N \cdot R_N$$

$$\frac{dMG^*}{dt} = \beta \cdot (1 - P_t)^{N+1} \cdot \left(1 - \frac{R_N}{K_{NC} + R_N}\right) - d_{mN} \cdot MG^* - r_b \cdot MG^*$$

$$\frac{dMG}{dt} = r_b \cdot MG^* - d_{MG} \cdot MG$$

Figure 3.3: Schematic (left) of the mechanistic steps of the sRNA transcriptional NAR network captured by the governing equations (right). Molecular implementations that highlight key rate constants and interactions are shown below each network schematic. N refers to the number of tandem repressors included in each construct. The equations model the tandem copies ($N = 2$) of repressor R_1 as a single new repressor, R_2. This effective model also used different degradation rate constants, d_{mN}, for each repressor species. Parameter and Species descriptions can be found in Table 1 and 2, respectively. The only difference between the negative autoregulation (NAR) and control (CTRL) constructs is the use of a mutated attenuator target region that disrupts repression, modeled as a change in repression coefficient K_{NC}, describing a weak non-cognate repression caused by crosstalk.

describe the maturation of this RNA reporter. Overall the inclusion of these maturation steps created two additional species to be modeled which expanded the number of ODEs to four.

In addition, we introduced terms in the model to account for both autotermination and cross-talk within the network. Previous work suggested that the presence of an upstream attenuator region reduces the amount of downstream transcription through

an autotermination mechanism that causes RNA polymerase to drop off even in the absence of repressor sRNA[11, 64]. This autotermination was modeled by including a single parameter that models the autotermination probability, Pt, based on our previous model[48]. In this work, we also found that transcription of repressor sRNA prior to the MG reporter also leads to a similar amount of reduction in MG fluorescence (Figure B.3A). This indicated that in general the transcription of upstream RNA species reduces the amount of downstream RNA that can be transcribed, likely due to general effects of the imperfect processivity of the RNA polymerase enzyme. Since both an upstream attenuator or repressor had similar autotermination effects, to incorporate this in our model we used the same parameter Pt to denote the probability of transcription being disrupted as polymerase transcribes one of the units of our NAR construct, which could be an attenuator or sRNA-ribozyme construct (Figure 3.3). Therefore, the probability of a particular sequence region being transcribed without disruption is modeled as $(1-P_t)$n where n is the number of upstream construct units (Figure 3.3).

We next incorporated the effect of cross-talk in our model. Experimentally, we sought to characterize the dynamics of the sRNA NAR network by comparing it to a control construct where only the negative feedback connection was broken with all other components unchanged. This can be achieved by mutating the attenuator sequence to an orthogonal sequence that does not respond to the repressor sRNA[64] (Figure B.1). While in principle this is a good control, the mutant pT181 attenuator is not completely orthogonal to the wild type repressor sRNA[64] (Figure 3.2B). We therefore had to account for cross-talk in order to accurately model the comparison between the full sRNA NAR network and the control network. Cross-talk repression for the control network was also modeled as a first order Hill function with a relatively large repression constant K_c.

Once our model was proposed, the next step was to determine all 12 unknown parameters (Table 3.2, Figure 3.3) that would provide a quantitative estimation of the time-dependent fluorescence trajectories generated by the NAR network in TX-TL. In previous work, we adapted a parameter sensitivity analysis method to guide the design of a series of TX-TL experiments to determine each unknown parameter[48]. For a given experiment modeled by a subset of model ODEs, sensitivity analysis determines which parameters cause the largest changes to the predicted behavior of the ODEs, i.e., which set of parameters the ODEs are most sensitive to. The most sensitive parameters can be found by computing the sensitivity matrix, which describes how the time-varying molecular concentrations of the genetic network in the experiment change in response to a change in the parameters of the model (see Methods). The most sensitive parameters for each experiment are then ideal candidates to extract by fitting the parameters to the experimental data[48]. Since multiple parameters can be estimated from a single experiment, it is therefore possible to estimate all relevant parameters from a reduced set of experiments that are each sensitive to different subsets of parameters.

We thus set out to perform sensitivity analysis on our NAR network. However, because of the feedback element in the NAR, the network is non-linear, which can result in coupled parameter sets[35], meaning that parameter values can be restrained by the values of others within the model. To avoid solving for coupled parameters, we proposed a set of experiments consisting of key components of the NAR network arranged in configurations that removed network non-linearity (Figure 3.4). Many of the constructs had already been characterized in previous work with SFGFP as the reporter instead of the MG aptamer[110, 48]. We therefore chose to continue to use SFGFP as the reporter for most of the initial parameterization experiments with one extra experiment to estimate key parameters related to the MG aptamer reporter construct. The use of SFGFP in these experiments however, resulted in four additional parameters to solve for to account for

protein translation, maturation and mRNA degradation (Figure 3.4, Table B.2). To perform the sensitivity analysis, we used an initial set of 16 parameter guesses taken from our previous work to calculate the sensitivity matrix from our proposed experimental design. We followed the guidelines from our previous study[48] and designed TX-TL experiments that could be performed at 29°C for 100 min with fluorescence monitored every 5 min. The sensitivity analysis on the proposed linear components from our NAR network suggested that we could identify all 16 parameters with five TX-TL experiments (Figure 3.4, Supplementary Notes B.1 and B.2).

We next sought to estimate the parameters in our model by performing the set of experiments designed through the sensitivity analysis. We performed replicate TX-TL experiments with each of the designed plasmid combinations and collected dynamic fluorescence trajectories of the reactions (Figure B.4). Using this experimental data, we then estimated parameters using an iterative fitting procedure that used each experiment in turn to find its designated parameters (see Methods). To capture the natural variation in experimental conditions, we initially input sets of parameters, drawn from uniform distributions centered around an initial best guess, into the fitting procedure. Each initial set was used as a starting point for the model, and was then optimized through the fitting procedure to give final parameter values (see Methods). This resulted in 10,000 sets of the 16 estimated parameters. To validate our method, we compared measured fluorescence trajectories for each experiment to simulated trajectories using 1,000 sets of parameters randomly sampled from the fitted parameter set distributions. This allowed us to compare the experiments to the mean simulated trajectory and 95% confidence intervals, which captured the range of trajectories predicted by our parameter distributions. As shown in Figure B.4, each mean simulated trajectory accurately described the experimental observation, with almost every measured experimental trajectory lying within the 95% confidence intervals of our simulation.

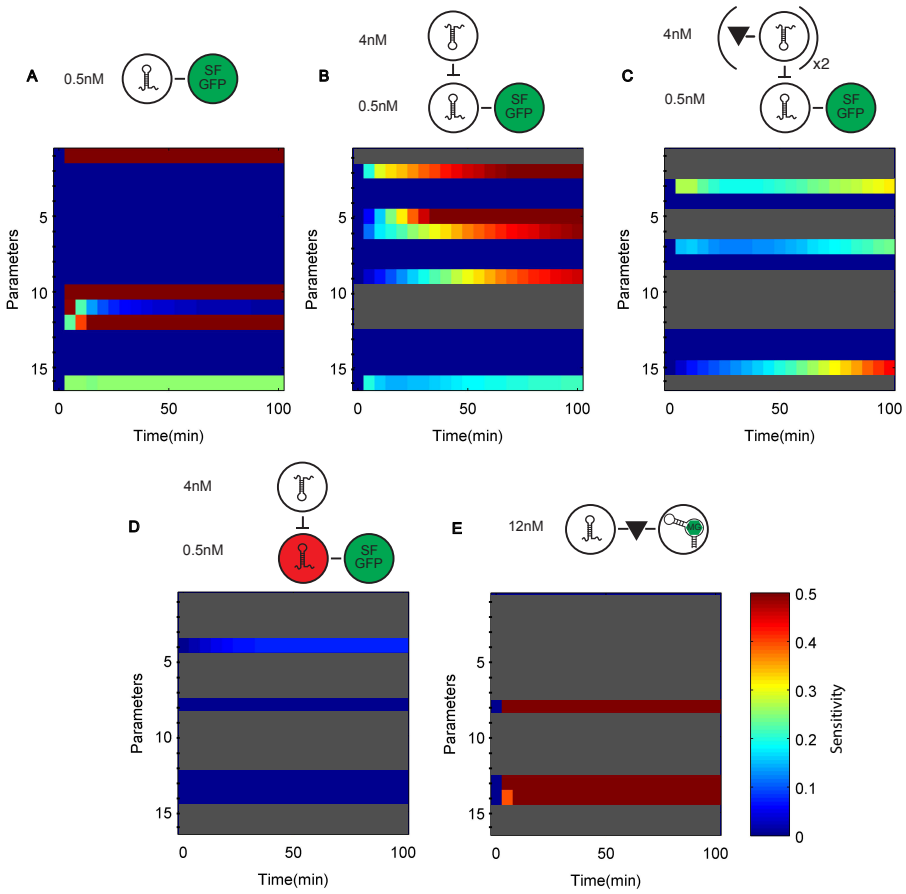


Figure 3.4: Sensitivity matrices for parameter identification experiments. Five parametrization experiments were designed on the basis of sensitivity analysis to estimate all 16 unknown parameters in our model (Supplementary Note Equations B.1.1-B.5.2). For each experiment, schematics of the constructs used are shown above the SFGFP portion of the calculated sensitivity matrix for that experiment, with DNA concentrations indicated. Parameters are numbered according to Table 3.2. Red/blue indicates high/low sensitivity, respectively. Note that time varying changes in parameter sensitivity indicates time windows of the trajectories that are influenced by each parameter. Experiments were designed in order from (A) to (E), with parameters identified by previous experiments marked as gray rows. (A) 0.5 nM of reporter plasmid alone is able to identify four parameters: β_m, k_t, α and d_m. (B) 0.5 nM of reporter and 4 nM of repressor plasmid is able to identify five parameters: K_1, d_1, r_{m1}, β_R and P_t. (C) 0.5 nM of reporter and 4 nM of plasmids containing repressor with two tandem repressors (separated by a ribozyme) is able to identify three parameters: K_2, r_{m2} and d_2. (D) 0.5 nM of reporter with a mutated PT181 attenuator and 4 nM of repressor plasmid is able to identify one parameter K_{C1}. (E) 12 nM of a plasmid with MG as reporter identifies the last three parameters: r_b, β and d_{MG}

Our goal of building a quantitative model was to accurately predict the behavior of the sRNA transcriptional NAR network. To assess the quality of the model prediction, we simulated dynamic trajectories of two-hour TX-TL experiments for no-autoregulation control, and NAR networks containing either single or tandem double repressor RNA configurations (Figure 3.5A). To test our model predictions, we next performed TX-TL experiments with the same constructs to obtain experimental fluorescence trajectories. For all four constructs, experimental results agree with the 95% confidence interval of our simulations (Figure 3.5B-C). We next sought to compute the response times of the experimental trajectories. To do this, we first found the simulated trajectory that had the most accurate fit to each replicate of the experimental data, and used these trajectories to estimate each response time (see Methods, Figure B.5). As suggested by the original qualitative model, the sRNA NAR network with a single sRNA did not demonstrate significant speedup of response time compared to its control network. In contrast, the sRNA network with tandem copies of repressor sRNA exhibited a clear separation of fluorescence trajectories compared to its control network, with a significant speedup of response time by four minutes with a p-value of 0.001 calculated using unpaired t-test[69] between the distributions of response times measured in each case.

Overall these results demonstrated that the sRNA transcriptional NAR network functioned in cell-free systems and extended the application of our effective modeling and parameterization methods to a non-linear sRNA network.

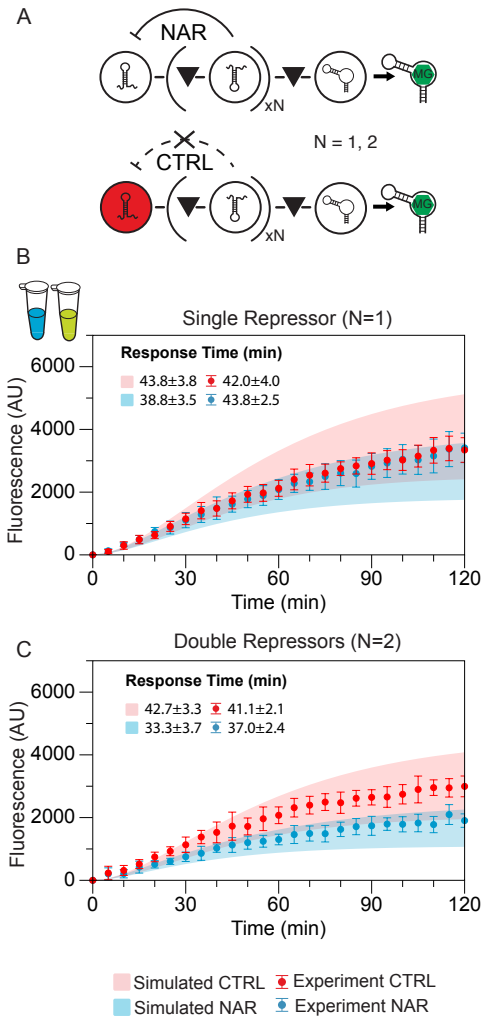


Figure 3.5: Model prediction and experimental characterization of the sRNA transcriptional NAR constructs in TX-TL. (A) Schematics of the NAR and control constructs designed for TX-TL using malachite green (MG) as a network reporter. (B) and (C) Comparison of experimental trajectories of MG fluorescence in TX-TL experiments (solid circles with error bars) with simulated model predictions (shaded regions) for single repressor NAR ($N = 1$) and double repressor NAR ($N = 2$) networks, respectively. Experimental error bars represent standard deviations of nine individual TX-TL reactions. Model simulated trajectories were generated by performing 1000 simulations with parameters randomly drawn from the set of 10,000 determined from the estimation procedure (see Methods). Predicted and measured response times for all four trajectories are shown in the upper left corner.

3.3.4 The sRNA transcriptional NAR network functions in *E. coli*

We next sought to adapt the NAR network for use in cells. Since the MG aptamer is difficult to express in cells[39], to accomplish this, we first replaced the MG reporter with a fast-degrading protein reporter, yem-GFP (Figure B.6)[104]. Initially, we tested the yem-GFP reporter in a simple context where the expression of yem-GFP was controlled by the pT181 attenuator in *E. coli* cells in the presence or absence of repressor sRNA. We found that the no-repressor GFP expression was lower than that of SFGFP, but the repression with repressor sRNA was nearly 100% (Figure B.6A). Next, in order to be able to assess cellular network dynamics from a defined starting time point, we replaced the constitutive promoter used above with the N-acyl homoserine lactone (AHL) inducible promoter Plux[15]. We compared the expression of the attenuator-yem-GFP construct under the control of the constitutive promoter and the P_{lux} promoter with varying concentrations of AHL and found that 100 nM of AHL resulted in maximum expression from the promoter (Figure B.6B).

Using these two modifications, we built NAR constructs analogous to those in Figure 6A and tested them in *E. coli* cultures while monitoring fluorescence every 20 minutes (Figure 3.6, B.7, B.8). As the model and TX-TL results suggested, the NAR construct with a single copy of repressor RNA did not show a significant difference in response time compared to its control construct *in vivo* (Figure B.8) due to inefficient repression strength of the single repressor when placed downstream of an attenuator element (Figure B.6C). However, experiments comparing the double repressor NAR network to the control showed that the yem-GFP trajectories began to reach steady state for the NAR network after approximately two hours while the control trajectories were still increasing. To calculate a response time for each construct, we performed regression analysis on data points to determine the steady state of individual fluorescence trajectories. We

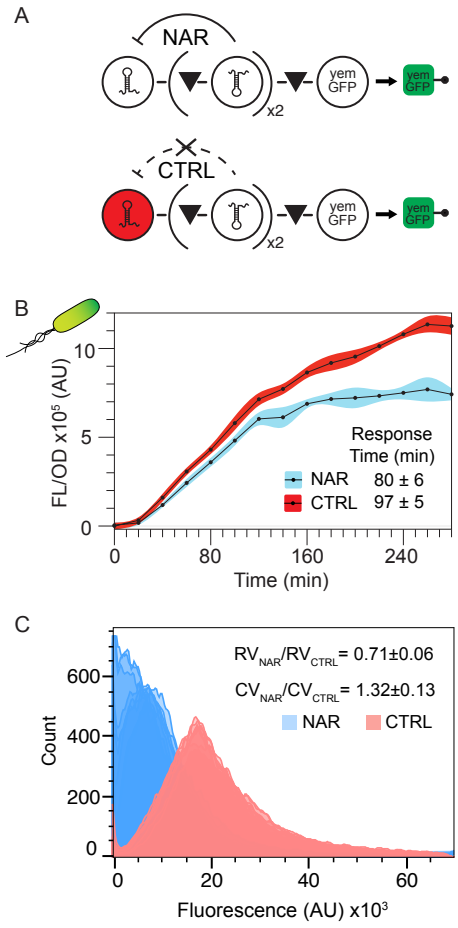


Figure 3.6: Demonstration and characterization of a functional sRNA transcriptional NAR network *in vivo*. (A) Schematics of the NAR and control constructs designed for *in vivo* testing using yem-GFP as a network reporter. (B) Time course of average fluorescence (FL/OD) collected from *E. coli* TG1 cells containing the double repressor NAR (blue) or control (red) construct over a five-hour period. Shaded regions represent standard deviations collected at each time point for nine biological replicates. Calculated response times (see Methods) indicate a functional NAR network that decreases the response time with respect to the control. (C) Flow cytometry histograms of a single time point measurement after 5 hours, tested with the double repressor NAR (blue) and control constructs (red) *in vivo*. Individual distributions are overlaid on each other with different shades of blue and red. The ratio of relative variance (RV) and coefficient of variation (CV) was calculated for each distribution and averaged. The averages of the ratio of (RV_{NAR}/RV_{CTRL}) and (CV_{NAR}/CV_{CTRL}) are shown in the upper right corner with errors representing the standard deviation of these calculated ratios.

then used the steady state level to find the time point at which each trajectory reached half of its steady state fluorescence value (see Methods). The response time ($T_{1/2}$) for the double repressor NAR network was 80±6 min, while the control showed a 97±5 min response time. This confirmed that the RNA-only NAR network functioned properly in *E. coli*.

3.3.5 Noise reduction properties of sRNA transcriptional NAR networks

In addition to reducing the steady state level and response time, NAR networks are also predicted to affect network noise properties. While the noise properties of protein-mediated NAR networks are well studied [99, 6, 27, 7, 97], little is known about the noise properties of sRNA transcriptional NAR networks. Our construction of the sRNA NAR network above thus gave us a unique vantage point to begin to study these properties.

The definition of network noise varies between studies, with most studies measuring noise with coefficient of variation (CV) of gene expression [7, 97, 6, 27], and others focusing on relative variance (RV), or Fano factor [118, 113]. While protein-mediated NAR networks have been demonstrated with a range of behaviors that either increase[45, 101] or decrease [7, 97] noise, under optimal conditions protein-mediated transcriptional NAR networks can reduce the CV. In particular, *Singh et al.* [100] predicted that to reduce noise measured by CV within a protein-mediated NAR network, there are different optimal windows of feedback strength needed for different orders of Hill function cooperativity in the regulatory interaction of the network. Specifically, they find that the CV of NAR network expression can decrease compared to the unregulated

networks only when the Hill coefficient of the repressive feedback is greater than one. Since our sRNA repression mechanism is believed to have a Hill coefficient of one[64], there is no indication that the sRNA transcriptional NAR network would reduce intrinsic noise measured in CV. Interestingly, sRNA mediated translational NAR networks are predicted to affect noise properties very differently[67, 118], and are generally not expected to reduce CV but may reduce RV.

To analyze the noise properties of our sRNA transcriptional NAR network, we performed a flow cytometry analysis of the double repressor NAR network and its control construct in *E. coli* for a specific time point at 5 hours after AHL induction (see Methods). We then calculated RV and CV from flow cytometry distributions to characterize network variation. All replicates showed a clear reduction in RV and an increase in CV (Figure 3.6C). These results align with previous studies on protein-mediated transcriptional NAR and sRNA mediated translational NAR networks. However, more careful noise studies are needed, such as experimental setups that allow the extraction of the intrinsic and extrinsic components of the total noise [30], before more precise conclusions can be drawn.

3.4 Conclusion

In conclusion, we demonstrated the utility of using effective models in combination with TX-TL experiments to systematically and rapidly prototype RNA network designs. Using this design and prototyping method, we were able to create the first functional sRNA mediated transcriptional NAR network that functions both in TX-TL and *in vivo*. We started the design process using a set of over-simplified ODEs to predict potential problems of initial NAR designs and to propose a design that used a double sRNA repressor

would be necessary to observe a reduction in response time that is characteristic of NAR networks. Using this model-guided design, we also significantly shortened the length of the construct that implemented this design, while preserving the proper functioning of all of the components. Similar to a recent investigation[124], we showed that the truncated version of the sRNA repressor is adequate for achieving the same repression strength of the full-length repressor. However, when a more compact transcriptional terminator is used, this truncated sRNA repressor RNA requires an extra piece of ribozyme between the repressor and the terminator to preserve orthogonality. This result offers an engineering guideline to concatenate multiple RNA components together for more complex sRNA networks.

Once our model had identified the overall NAR design, we then adapted our previously developed sensitivity analysis based parameterization method to construct and parameterize a more detailed model of the network. Using this approach, we designed five simple TX-TL experiments that we used to parameterize all unknown parameters. These parameters were then used to quantitatively predict the dynamic trajectories of two versions of the sRNA transcriptional NAR network. The model prediction suggested that the double repressor version would function according to the two key characteristics of NAR networks: faster response time and reduced steady-state level. TX-TL characterization of the network revealed that the model predictions were both qualitatively and quantitatively accurate. Finally, we modified the TX-TL-prototyped NAR constructs for *in vivo* experiments and showed that the network also demonstrated faster response time and lower noise in *E. coli*. Interestingly, the reduction in noise followed a similar trend for our sRNA mediated transcriptional NAR network as that observed for protein-mediated translational NARs and sRNA mediated translational NAR networks[118].

Several aspects of this work are significant. First, this work represents the first tran-

scriptional sRNA negative auto-regulatory circuit that functions both in cell-free reactions and in living cells. Second, in this work, we extended our parameterization procedure to a non-linear network. Because of the natural feedback component of NAR networks, it is difficult to parameterize its parameters independently. By parameterizing components individually, we were able to incorporate the feedback non-linearity at a later modeling step which we showed was able to recapitulate experimental results. Finally, this work represents a streamlined and quantitative model guided design and TX-TL based prototyping methodology for in-cell RNA network engineering. We found that the parameters we obtained from TX-TL parameterization procedures are adequate to predict the function of our sRNA transcriptional NAR network accurately and provide insights for TX-TL prototyping experiments that then inform the construction of networks that function in cells. We anticipate this design-build-prototype-test network building process will minimize cloning effort and speed up network development cycles as the field moves towards more sophisticated synthetic network engineering.

3.5 Method

3.5.1 Plasmid construction and purification

A table of all the plasmids used in this study can be found in Supporting Information Table S=B.5, with key sequences found in Table B.4. The pT181 attenuator and repressor plasmids, pT181 mutant attenuator and repressor plasmids were constructs pAPA1272, pAPA1256, pAPA1273, and pAPA1257 respectively, from Lucks et al.[64]. All NAR and control constructs were created using Golden Gate[32] assembly and Gibson assembly. The NAR constructs were created with pT181 attenuator and pT181 repressor,

the control constructs were created with pT181 mutant attenuator and pT181 repressor. Plasmids were purified using a Qiagen QIAfilter Plasmid Midi Kit (Catalog number: 12243) followed by isopropanol precipitation and eluted with double distilled water.

3.5.2 TX-TL experiments

TX-TL buffer and extract was prepared according to Garamella et al[36]. TX-TL buffer and extract tubes were thawed on ice for approximately 20 min prior to performing experiments. Separate reaction tubes were prepared with combinations of DNA representing a given network condition. Appropriate volumes of DNA, buffer, and extract were calculated using a custom spreadsheet developed by Sun et al[106]. Buffer and extract were mixed together and incubated for 20 min at $37°C$. DNA for the specific experiment was then added into each tube according to the previously published protocol[109]. Ten μL of each TX-TL reaction mixture was transferred to a 384-well plate (Nunc 142761), covered with a plate seal (Nunc 232701), and placed on a Biotek SynergyH1m plate reader. Temperature was controlled at $29°C$. Superfolder GFP (SFGFP) (485 nm excitation, 520 nm emission, 100 gain) or malachite green (MG) (610 nm excitation, 650 nm emission, 150 gain) fluorescence was measured every 5 min for 120 min. MG constructs were observed to immediately generate observable signal in between mixing and loading the plate reader, making quantification of early time points difficult. Therefore, the first measurements of all constructs with MG were assumed to be 0 and data collection was timed to start 5 minutes after reactions were mixed. Each reaction was run with a minimum of triplicate technical repeats, and repeated three times independently (minimum of nine total replicates). A ten μL sample of TX-TL buffer, extract, and water was run together with each independent experiment as blank. All data points were then processed with blank values subtracted at each time point.

3.5.3 Network/Model Simulations

All models were simulated by solving corresponding ODEs using *Matlab* function $ode15s$ over a set of discrete time steps using guessed or estimated parameters. All initial conditions for species concentrations were zero.

3.5.4 Parameterization experiment design

All unknown parameters were estimated by fitting model simulations to a small set of time course data generated from TX-TL experiments. Following previous work[48], these experiments were designed through an iterative process of identifiability tests on the set of model equations (Figure 3.4, Supplementary Note B.1). An individual experiment was expected to produce a measurable trajectory of fluorescence as a function of time. After a specific experiment was proposed, we determined which parameters were identifiable from this experiment through sensitivity analysis (See Supplementary Note B.2). The iterative experiment starts with the simplest experiment that has the smallest number of network components and thus the smallest number of parameters governing the fluorescence trajectory. Identifiable parameters from this experiment are then estimated by fitting models to the experimental data as described below (Parameter estimation). The next experiment that contains a greater number of components and the analysis repeated, except that parameters already identified by previous experiments were marked as determined and were not included in sensitivity analysis. Rounds of experimental design and sensitivity analysis were performed until all 16 parameters could be identified by 5 TX-TL experiments (Figure 3.4).

3.5.5 Parameter estimation

Parameters were estimated from each designed experiment by fitting the identifiable parameters of that experiment to measured SFGFP expression time trajectories. The parameter estimation problem is given by:

$$min(G_{t_k}^{experimental,mean} - G_{t_k}^{simulated})^2$$

Where $G_{t_k}^{experimental,mean}$ denotes the average value of the experimentally observed SFGFP expression at a certain time t_k. The vector P contains all of the identifiable parameters in the experiment being analyzed. $G_{t_k}^{simulated}(P)$ denotes the model simulated SFGFP expression at time t_k.

To optimize the initial guesses, we first started with a set of initial guess based on pervious work (Table B.1)[48]. We then constructed uniform distributions (100 points) around each parameters value (15%), generating 100 sets of parameters, which led to 100 simulated dynamic trajectories for each of the 5 estimation experiments designed by sensitivity analysis procedure. We picked the set of parameters that generated the best fit to the 5 sets of experimental data collectively, and used it as a starting point for the next round. We repeated this procedure 10 times to achieve the best guess. We then started the estimation process with the best-guessed parameter set and constructed uniform distributions (10000 points) around each parameter value (15%). Each set of parameters served as a different starting point for finding optimal parameter estimates. Each optimal set was found by sequentially applying the fitting procedure to each of the designed parameters using the *Matlab* function $fmincon$, only fitting the identifiable parameters for that experiment. Fit parameters from one experiment were then used to update and replace the initial guesses before moving on to the next experiment until each of the 16 unknown parameters were fit from the 5 experiments. The result of this procedure produces 10000 sets of estimated parameters, which were then subject to the

analysis described in the main text.

3.5.6 Strains, Growth Media and *in vivo* Gene Expression Experiments

All experiments were performed in *E. coli* strain TG1. Plasmids with different network constructs were transformed into TG1 competent cells, plated on LB + Agar plates containing 100 μg/ml carbenicillin, and incubated overnight at 37°C. At least three colonies of each experimental condition were inoculated into 300L of LB containing carbenicillin in a 2mL 96-well block (Costar 3960), and grown for 12 hours overnight at 37°C at 1000 rpm in a Labnet Vortemp 56 bench top shaker. Four microliters of the overnight culture were added to 196 l of M9 supplemented M9 media (1X M9 minimal salts, 1 mM thiamine hydrochloride, 0.4% glycerol, 0.2% casamino acids, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$) containing carbenicillin and grown for 4 hours at 37°C at 1000 rpm.

For time course measurements, the sub-culture was then diluted to 0.015 OD with M9 + carbenicillin and 100nM N-Acyl homoserine lactone (AHLs) to a total volume of 1mL and grown at 37°C at 1000 rpm. During the 5-hour time course, 50μL of culture was removed from of the main culture every 20 minutes for measurements. Fluorescence (485 nm excitation, 520 nm emission) and optical density (OD, 600 nm) were measured on a Biotek SynergyH1m plate reader. For steady state expression measurement, the sub-culture was then diluted to 0.015 OD with M9 + carbenicillin and 100nM N-Acyl homoserine lactones (AHLs) to a total volume of 200μl and grown at 37°C at 1000 rpm for 5 hours. Fluorescence (485 nm excitation, 520 nm emission) and optical density (OD, 600 nm) were then measured on a Biotek SynergyH1m plate reader.

For flow cytometry measurement, the sub-culture was then diluted to 0.015 OD with M9 + carbenicillin and 100nM N-Acyl homoserine lactones (AHLs) to a total volume of 3ml and grown in culture tubes for 5 hours at 37°C at 220rpm. A total of 50,000 events were collected on a BD AccuriTM C6 plus Flow Cytometer during measurement.

All experiments contained three replicates from at least three different transformed colonies, and were repeated independently three times for a minimum of nine total replicates.

3.5.7 Response-time Calculation

To estimate the response time $T_{1/2}$ of a network, we estimated the time for individual dynamic trajectories to reach steady state based on the rate production of fluorescence signal. We first scaled all data points by the maximum experimentally observed fluorescence value across all four experiments F_{max}^{expr}:

$$\hat{F}_{t_k}^{expr} = \frac{F_{t_k}^{expr}}{F_{max}^{expr}}$$

: $F_{t_k}^{expr}$ is the experimentally observed fluorescence/OD value (malachite green for in vitro networks and yem-GFP for in vivo networks) at any time during the experiment. For in vitro data sets, where the variability in fluorescence values could complicate data analysis, we created a pool of 100 simulated trajectories for each experimental trajectory from 1,000 sets of randomly sampled parameters. We then compared the R_2 values for each of the 100 simulated trajectories and the experimental trajectory, and selected the simulated trajectory that gave the max R_2 value for response time analysis (Figure 3.5, B.5).

To calculate trajectory response times, we first used least squares to estimate the rate of SFGFP signal production at each time point in order to find the steady-state value of

each trajectory, which was defined as the trajectory value when the rate of production drops below a threshold value. Beginning with the first normalized data point ($k = 1$) in the data set, we took a total of five data points ($n = 5$) and calculated the least squares slope (m) at time point k with following equation[123].

$$m_{t_k} = \frac{n\sum_{i=k}^{k+n-1}(t_i \cdot \hat{F}_i^{expr}) - \sum_{i=k}^{k+n-1} t_i \cdot \sum_{i=k}^{k+n-1} \hat{F}_i^{expr}}{n\sum_{i=k}^{k+n-1}(t_i \cdot t_i) - \sum_{i=k}^{k+n-1} t_i \cdot \sum_{i=k}^{k+n-1} t_i} \tag{3.1}$$

The objective of least squares is to find slope (m) that minimizes the sum of squared errors (r_2) of the differences between the observedSFGFP production and those predicted by a linear model. The m found became the slope at time point $t_k = 1$, and we stepped into the next set of five data points with $k = 2, 3, ,$ etc., creating an array of slopes to describe the fluorescence trajectory.

When the absolute value of the calculated slope is below a set threshold of 0.005, the trajectory is defined to be at steady state and its fluorescence level at steady state is recorded. For trajectories that did not reach steady state at the end of the experiment, we forecasted its trajectories beyond the experimental time duration and predicted the time of steady state, assuming the SFGFP signal production rate declines at a constant rate. Using the steady state signal values, we traced through the trajectory to identify the response time when the network signal reached half of its steady state value. If the exact value was in between two experimental data points, a linear interpolation between the two experimental data points was used to estimate the response time.

Using the steady state signal values, we traced through the trajectory to identify the response time when the network signal reached half of its steady state value. If the exact value was in between two experimental data points, a linear interpolation between the two experimental data points was used to estimate the response time.

3.5.8 Signal Noise Analysis

Noise magnitude was considered with two measures: the coefficient of variation (CV) of signal X and the relative variance (RV) or Fano factor of signal X. They were calculated with following formulas:

$$CV_x = \frac{\sigma(X)}{\mu(X)} \tag{3.2}$$

$$RV_x = \frac{\sigma(X)^2}{\mu(x)} \tag{3.3}$$

Where $\sigma(X)$ is the standard deviation of signal X and $\mu(X)$ is the average value of signal X. Mathematically, CV of signal X can increase when the mean decreases while RV does not decrease sufficiently[118].

3.6 Acknowledgement

The authors gratefully acknowledge the gift of TX-TL extract and buffers used in this work from Vincent Noireauxs laboratory (University of Minnesota). The authors also thank Marshall Colville for early help in characterizing NAR circuitry, and Adam Silverman for helpful comments on this manuscript.

CHAPTER 4

UNCOVER THE DESIGN PRINCIPLE OF GENERIC TRANSFER FUNCTION MANIPULATION UTILIZING SMALL RNA REGULATORS [1]

4.1 Abstract

Inducible promoter systems are at the heart of many applications in molecular biology. While there is a range of potentially useful inducible promoter systems, their transfer functions—the relationship between molecular input and molecular output—are static and often require tuning via molecular engineering. In previous work, we showed that Small Transcription Activating RNAs (STARs) can be fused to *Lux* operon to form a functional STAR hybrid operon; and these hybrid operons have different transfer functions when built with different STARs. In this work, we combined theory, simulation and experiments to uncover the factors that govern this transfer function shift. With this interdisciplinary approach, we also discovered the universal dynamic range amplification phenomenon induced by STARs and uncovered the underlying mechanistic causes of this phenomenon. We showed that this design is generic and extendable to other inducible systems but it has limitations. We concluded with a design principle of STAR-induced dynamic range amplification that would guide future designs of this application. We envision that this work will contribute to our understanding to biological systems and improve our ability to engineer synthetic circuitry.

[1] THIS WORK IS BEING PREPARED AS A MANUSCRIPT FOR REVIEW.

Introduction

Transcriptional regulatory networks are essential elements of living cells. They are tasked the responsibility of communicating between the cells and their environments. Cells rely on these networks to sense many signals, including temperature, light, the presence of other organisms, nutrient and harmful toxins. The ability to sense environmental changes and response to these changes is crucial to the survival of cells. These transcriptional regulators were developed by nature through millions of years of evolution. Since the first discovery of transcriptional regulator the *Lac* operon by Jacob and Monod[55], scientist have been studying and isolating transcriptional regulators and repurposing them for other tasks. These studies eventually lead to the birth of synthetic biology, which is a study with a major purpose of designing and engineering synthetic gene networks to reprogram living cells. Transcriptional regulators are the fundamental building blocks of synthetic circuitry. Most of them are transcriptional factor based like the classic *Lac*, *lux* and Tet operons; some of them are RNA based like riboswitches[95] and antisense RNAs[64, 20], some of them are in the category of two component systems[103] like the light sensor[83]. Regardless their specific mechanism, they are inducible systems and their performances are all defined by their transfer functions, which is an empirical measurement that describes how the output changes as a function of the input[117]. Transfer functions are defined by the molecular interactions among parts involved in the inducible system. Thus shifting transfer functions often requires molecular engineering, which is a nontrivial task. Therefore, engineering well-performing inducible systems has remained one of the major focuses in synthetic biology.

Among all inducible systems, molecular interactions between small RNAs are relatively more discernible and easier to engineer. Regulatory RNAs have emerged as

powerful tools in synthetic biology in recent years. Antisense sRNA as a part of this expanding tool box has been studied extensively. A growing body of work focuses on uncovering the design principles of regulatory RNAs, making use of a powerful suite of computational[114, 130] and experimental[63, 122, 105, 111] methods. These works have improved our ability to engineer RNAs and made regulatory sRNA highly designable. As a result, we have expanded the tool box of antisense regulatory RNA drastically[110, 124, 20, 22]. In recent studies, we created small transcriptional activator RNAs (STARs) by inverting the attenuation mechanism of the natural occurring transcriptional repressor RNAs[20]. We showed that STARs can be fused with transcriptional factor mediated activators *lux* to form a STAR hybrid inducible operon (Figure 1A). Interestingly, we observed drastic transfer function shift when we compared 3 different versions of hybrid operons that were built with 3 different STARs (Figure 1B). This was an intriguing observation, but the mechanism behind this phenomenon has remained unknown.

In this work, we adapted an interdisciplinary approach that combines theory, simulation and experiments to study this observation. From our previous computational work on regulatory repressor RNAs, we learned that, first, they can be modeled with effective ordinary differential equations (ODEs); second, the parameters can be estimated easily with the sensitivity analysis based parameterization method[48]. Based on these findings, we isolated and estimated the governing parameters that caused the transfer function shift. As a result, these estimated parameters recreated transfer functions that largely resemble our previous observation. Next, we combined theory, simulation and experiments to show that this hybrid operon design offers an insulated, structural independent transcriptional tuning knob to tune down the signal leakage of an inducible system. At the same time, we showed that when this tuning knob is tuned up, a dynamic range (ON state signal/ OFF state signal) amplification is robustly observed. We

then uncovered the underlying causes of this amplification effect and showed that it is universal with STARs variants. In addition, we showed that the STAR sequester would also impact the transfer function and dynamic range of the system when combined with the hybrid operon. Finally, we applied this design to other systems to show its genericity and extendibility. We concluded this work with a discussion on the limitation of this amplification effect, and established a design principle to guide future application of STARs as a dynamic range amplifier.

4.2 Results and Discussion

4.2.1 Understand the transfer functions of STAR hybrid activators

In our past work on STARs lead by Chappell et al [22], we showed that STARs can serve as an intermediate activator between an inducible promoter and the reporter gene, as shown in figure 1A. The inducible promoter, which is regulated by a transcriptional factor (TF) protein, controls the transcription of STAR molecules. The reporter gene is driven by a constitutive promoter followed by Target. The Target is a short RNA piece that folds into a termination hairpin to prevent transcription of the reporter mRNA when STAR is absent. When STAR is present, the interaction with target would open up the termination hairpin, therefore, allows transcription of its downstream gene. In previous work, we tested three versions of the same design with three different STAR/ Target pairs fusing with the Lux operon. We found that these three hybrid operons have very different transfer functions and dynamic ranges (Figure 4.1B)[22]. This observation was surprising considering all STARs were designed with very similar mechanistic functions and lengths. Transfer functions define the fundamental characteristics of inducible systems and determine their potential applications in synthetic biology. They are usually dictated by the biomolecular interactions between molecules that are involved in the induction mechanism. Therefore manipulating transfer functions is not a trivial task. Intrigued by the observed transfer function shifts caused by STARs, we set off to gain a fundamental understanding of this observation and discover the principle of manipulating transfer functions and dynamic ranges with STARs. We started with writing simple

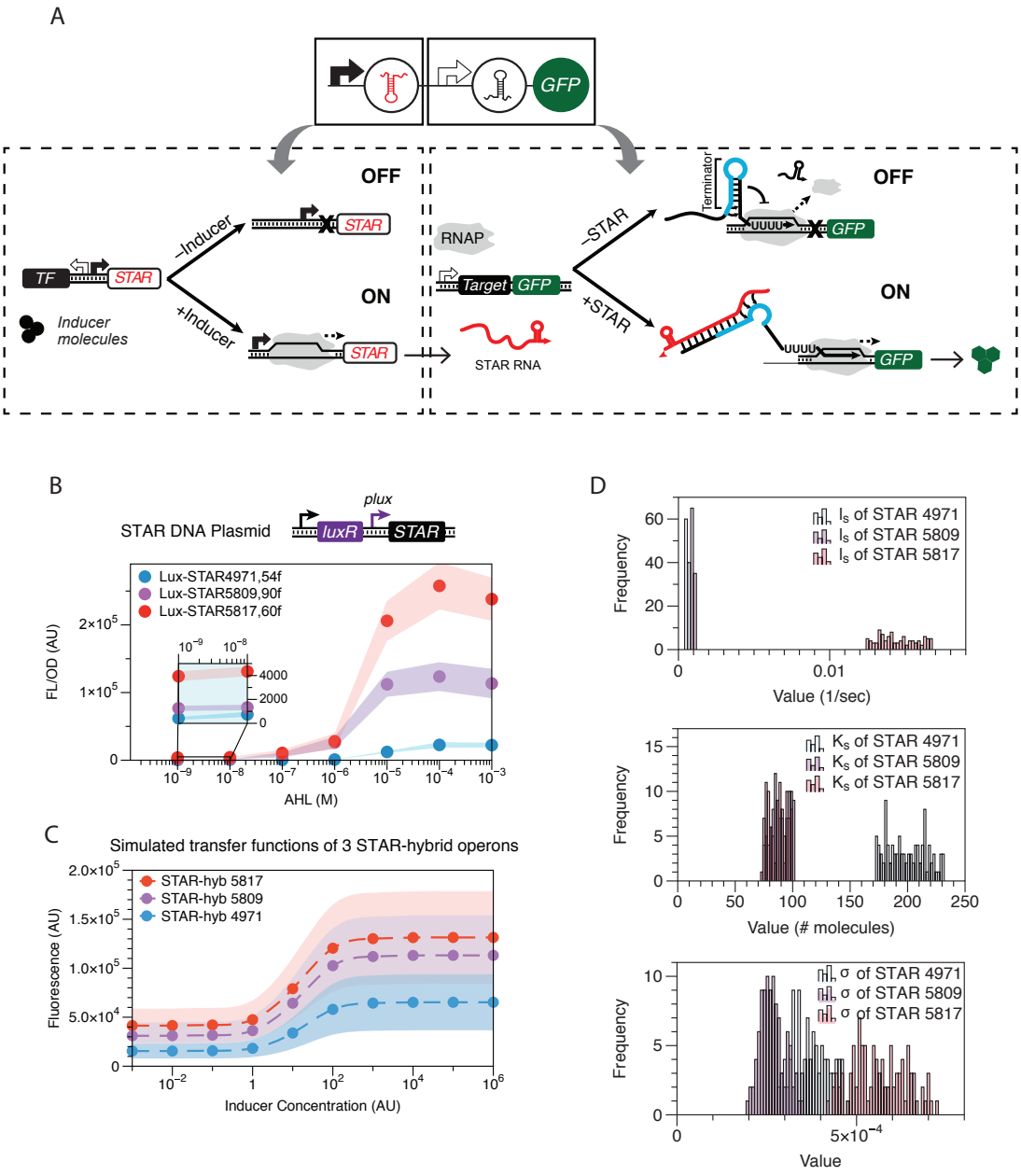


Figure 4.1: Understanding the transfer function shift caused by STARs. (A) The schematics of the STAR hybrid operon. The inducible promoter regulates the expression of STARs; STARs then regulates the expression of GFP. (B) Re-plot of an observation from previous work[22]. Three variants of STAR hybrid inducible operon presented 3 different transfer functions. (C) Transfer functions simulated based on estimated key STAR parameters. They largely resemble the observed transfer function from B. (D)Key STAR parameters estimated based with STAR circuit dynamics in TX-TL, see Appendix C1

Table 4.1: Species involved in hybrid inducible system

Parameter	Description
β_m	Transcriptional rate of RNA driven by constitutive promoter
β	Transcriptional rate of RNA driven by inducible promoter P_x
x	Concentration of inducer in system
K_x	Activation coefficient of inducer x
l_x	Leak coefficient of inducible promoter ranges $0 - 1.0$
d_m	Degradation rate of mRNA
k_t	Translational rate of GFP
d_p	Degradation rate of GFP
K_S	Activation coefficient of STAR
l_S	Leak coefficient of STAR ranges $0 - 1.0$

effective ODEs to describe the molecular dynamics of the hybrid operon in Figure 4.1.

$$\frac{dS}{dt} = \beta \cdot \left(\frac{x}{K_x + x} + l_x\right) - d_x \cdot S \tag{4.1}$$

$$\frac{dM}{dt} = \beta_M \cdot \left(\frac{S}{K_S + S} + l_S\right) - d_M \cdot M \tag{4.2}$$

$$\frac{dP}{dt} = k_t \cdot M - d_p \cdot P \tag{4.3}$$

All parameters and species are listed in Table 4.1 and Table 4.2:

Table 4.2: Species involved in hybrid inducible system

Species	Description
S	STAR
M	mRNA of GFP
P	Reporter protein

When $x = 0$, assuming steady state:

$$\frac{dS}{dt} = \beta \cdot l_x - d_S \cdot S = 0 \tag{4.4}$$

$$\frac{dM}{dt} = \beta_m \cdot \left(\frac{S}{K_S + S} + l_S\right) - d_m \cdot M = 0 \tag{4.5}$$

$$\frac{dP}{dt} = k_t \cdot M - d_P \cdot P = 0 \tag{4.6}$$

Solving the above equations, we obtain the following expressions with $\sigma = d_S/\beta$

$$S = \frac{l_X}{\sigma} \tag{4.7}$$

$$M = \frac{\beta_m}{d_m} \cdot \left(\frac{l_x}{\sigma K_S + l_x} + l_S\right) \tag{4.8}$$

$$P_{hyb,off} = \frac{k_t \beta_m}{d_P d_m} \cdot \left(\frac{l_x}{\sigma K_S + l_x} + l_S\right) \tag{4.9}$$

When $x \to \infty$, assuming steady state

$$\frac{dS}{dt} = \beta \cdot (1 + l_x) - d_S \cdot S = 0 \tag{4.10}$$

$$\frac{dM}{dt} = \beta_m \cdot \left(\frac{S}{K_S + S} + l_S\right) - d_m \cdot m = 0 \tag{4.11}$$

$$\frac{dP}{dt} = k_t \cdot M - d_P \cdot P = 0 \tag{4.12}$$

Solving the above equations, we obtain the following expressions with $\sigma = d_S/\beta$

$$S = \frac{1 + l_x}{\sigma} \tag{4.13}$$

$$M = \frac{\beta_m}{d_m} \cdot \left(\frac{1 + l_x}{\sigma K_S + 1 + l_x} + l_S\right) \tag{4.14}$$

$$P_{hyb,on} = \frac{k_t \beta_m}{d_P d_m} \cdot \left(\frac{1 + l_x}{\sigma K_S + 1 + l_x} + l_S\right) \tag{4.15}$$

In 4.9 and 4.15, the first group of terms $\frac{k_t \beta_m}{d_P d_m}$ are all determined by the transcriptional and translational characteristics of the reporter protein. The leak coefficient from the original inducible operon l_x is defined by the molecular characteristics of the operon. These parameters are not impacted by the presence of STARs. Therefore, the terms that

describe the impact of STAR on the on and off state of gene expression and the transfer function of the operon are σ, K_S ,and l_S . In which $\sigma = d_s/\beta$ describes the turnover rate of STARs; K_S and l_S are the activation constant and leak constant of STARs, respectively. As a result, we hypothesized that the changes in the response curves were causes by some, or all of these parameters of STARs.

We have learned in our past work that different repressor sRNAs would have different parameters associated with their own regulatory mechanistics [48, 46]. Therefore we hypothesized that STARs would have similar properties. We adapted our sensitivity based paramerization procedure for tree pairs of STARs that were used in the previous study (See Appendix C)[48]. The estimated parameters agreed with our previous observation of these STAR pairs (Figure 4.1D)[20]. For instance, we know that STAR 4971 has the lowest off state leak in vivo but also the lowest on state signal. It has the lowest leak coefficient l_S and the highest dissociation coefficient K_s in the group. STAR 5817 showed the highest on state expression but also the highest leak, this observation also aligns with the estimated high leak coefficient l_x and its low activating constant K_s. Although these parameters were not estimated in living cells, we found in previous work that it serves a prototyping tool for circuit performance *in vivo*. We then simulated Equation 4.1 to 4.3 with STAR parameters drawn from the estimation pool, in combination with some parameters from literature (Appendix C Section 1). As a result, we found that our simulation agreed with the observation that each hybrid operon has a distinct transfer function (Figure 4.1C). Specially, the STAR hybrid 4971 yields a transfer function that has the lowest off state leak and lowest on state signal. STAR hybrid 5809 is in the middle, where STAR hybrid 5817 has the highest on and off state values. This simulated result largely resembles our experimental observation from our previous work (Figure 4.1A). The intriguing observation of transfer function shifting caused by STARs led us to the quest of understanding the mechanism behind it. We started with theoretical

derivation to identify the parameters that dictate the the transfer function of the STAR hybrid operons. Then we estimated these parameters of STAR for the first time using our sensitivity based parameterization method. The resulting parameters not only align with our knowledge of these three pairs of STARs, but also simulated transfer functions for these 3 pairs of STAR hybrid operons that qualitatively agree with the observation of transfer function shifting.

4.2.2 STARs equips transciprtional inducible systems with a tuning knob

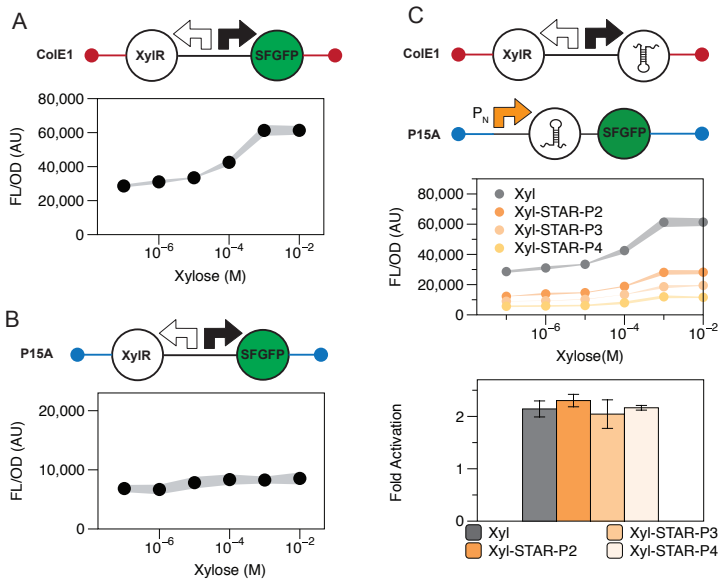


Figure 4.2: STAR hybrid operon provides an tuning knob to manipulate the otherwise static transfer function. (A) The transfer function of a *Xylose* operon on high copy ColE1 backbone, with a two fold dynamic range. (B) The entire transfer function moved to a medium copy backbone P15A. The operon became non-inducible. (C) The STAR hybrid operon leaves a constitutive promoter open for substitution (orange). Three weaker Anderson promoters were titrated in the configuration, the transfer function gradually moves down, reduces signal leak of the operon.(D) The dynamic ranges of these four transfer functions stay relatively constant

As we gain a better theoretical understanding on the impact of STARs on the transfer

functions, we began to explore the potential applications of STARs on transfer function manipulation. We took the first step to identify the challenge of transfer function manipulation. In order to understand what is dictating the dynamic range and transfer function of a inducible system, we write simple equations to describe the gene expression event of a protein P controlled by an activating transcriptional regulator that responses to inducer x.

$$\frac{dM}{dt} = \beta \cdot (\frac{x}{K_x + x} + l_x) - d_m \cdot M \tag{4.16}$$

$$\frac{dP}{dt} = k_t \cdot M - d_p \cdot P \tag{4.17}$$

Equation 4.16 and 4.17 represent the transcriptional and translational events in gene expression. Here M denotes the concentration of mRNA, and P denotes for the concentration of protein. β and k_t denote the rate of transcription and translation, respectively. d_m and d_p denote the degradation rate of mRNA and protein, respectively. K_x is the activation strength and l_x represents the leak, which ranges from 0 to 1. It defines the amount of mRNA being expressed when the system is not being induced. We assume here that there is no cooperative behavior of induction for simplicity.

If we assume steady state and rearrange the equations:

when $x = 0$,

$$P_{off} = \frac{k_t \beta}{d_p d_m} \cdot l_x \tag{4.18}$$

when $x \to \infty$,

$$P_{on} = \frac{k_t \beta}{d_p d_m} \cdot (1 + l_x) \tag{4.19}$$

Therefore, the signal gain of this induction is

$$f_{gain} = P_{on} - P_{off} = \frac{k_t \beta}{d_p d_m} \tag{4.20}$$

and the dynamic range of this induction is

$$f_{DR} = \frac{P_{on}}{P_{off}} = 1 + \frac{1}{l_x} \tag{4.21}$$

According to equation 4.20 above, signal gain of this transnational activator is governed by transcription rate β, translation rate k_t, mRNA degradation rate d_m and protein degradation rate d_p. However, none of these parameters would dictate the dynamic range of this activator. The only parameter that impacts the dynamic range is $l_x$constant , the leak, according to equation 4.21. The causes of leak differ from systems to systems. We explore the determine factors of leak coefficient l_x on several common inducible systems in supplemental information. (Appendix C section 3)

Inducible systems are often integrated into complex networks to aid the performance of various tasks. Depending on the task and the context of the network, these inducible systems often have to be tuned accordingly. Although it is ideal to increase the dynamic range of the induction f_{DR}, we learned from 4.21 that l_x is the sole parameter that governs dynamic range. The leak l_x is a difficult parameter to tune (see Appendix C). Here we focus on tuning the background signal P_{off} and P_{on}. Equation 4.19 and equation 4.18 show that these two factors are governed by four other parameters: transcription rate β, translation ratek_t, mRNA degradation rate d_m and protein degradation rate d_p. In which only β and d_m dictate the transcriptional level.

To design experiments for parameter tuning, there are limitations. First of all, although it is fairly straightforward to tune the signal up and down with RBS tuning, this effect is only translational. This would limit the operon's application to only directly regulate protein expressions. If one were to tune the signal gain at a transcriptional level, the task becomes challenging. For instance, transcription rate β is mostly tuned through promoter strength. However inducible promoters usually have fixed transcriptional strengths defined by its sequence. One other to to tune β is moving the promoter to another plasmid backbone with a different copy number. However the number of suitable plasmid candidates are limited. It could also result in a non-functional inducible

system due to over-tuning of promoter strength. To tune the other parameter d_m, a stability hairpin could be added to the upstream of mRNA in order to decrease d_m through mRNA stabilization. However, there's lack of biomolecular tools that destabilize mRNA

.

To test our theoretical understanding of tuning, we experimentally attempted to tune β and d_m using a xylose operon activator from iGEM registry. This xylose operon is inducible by the sugar D-xylose. Originally it has about two fold activation on the backbone ColE1 (Figure 4.2A). As we moved the entire operon to a lower copy plasmid P15A, the operon became barely inducible (Figure 4.2B). We also added a stabilization hairpin to the mRNA to tune d_m, this did amplify both the leak and the on level signal, with the dynamic range staying relatively constant (Figure 4.1C).

Since we found that STARs can alter transfer functions (Figure 4.1), we set off to understand what design criteria has to be met when fusing an operon with STARs in order to maximize its functionality. Here we continue with mathematical derivation to investigate the impact of STAR intermediate on the leaky signal at off state.

We know the leak of the original operon in equation 4.18 and the leak of the hybrid operon in equation 4.9. If we compare the leak between these two operons:

$$\frac{P_{hyb,off}}{P_{off}} = \frac{\frac{k_t\beta_m}{d_P d_m} \cdot \left(\frac{l_X}{\alpha K_S + l_x} + l_S\right)}{\frac{k_t\beta}{d_P d_m} \cdot l_x} \tag{4.22}$$

Here we introduce a new term $\theta = \beta_m/\beta$, equation 4.22 simplifies to:

$$\frac{P_{hyb,off}}{P_{off}} = \theta \cdot \left(\frac{1}{\sigma K_S + l_x} + \frac{l_S}{l_x}\right) \tag{4.23}$$

Leak is reduced when equation 4.23 is less than 1. To optimize the leak reduction, we want to minimize this equation. Therefore, theoretically, these criteria of th design would favor leak reduction. 1. Maximize σ: fast turn-over rate of RNA intermediate. 2. Small l_S : less leaky intermediate. 3. Large K_S: relatively weak activation by interme-

diate. 4. Small θ: Weak transcriptional rate of reporter mRNA.

However, this is just one of the two ends of the operons functionality, to ensure the basic inducible function of this operon, we need to look at its signal gain. The absolute signal gain f_{gain}:

$$f_{hyb,gain} = G_{hyb,on} - G_{hyb,off} \tag{4.24}$$

$$= \frac{k_t \beta_m}{d_P d_m} \cdot \left(\frac{1 + l_x}{\sigma K_S + 1 + l_x} - \frac{l_x}{\sigma K_S + l_x} \right) \tag{4.25}$$

This equation has a climax in the space of σ and K_S. We maximized $f_{Hyb,fold}$ in the direction of σ and K_S by taking partial derivatives with respect to these parameters and set them to 0.

$$\frac{\partial f_{hyb,gain}}{\partial \sigma} = \frac{k_t \beta_m}{d_P d_m} \left(\frac{K_S l_x}{(l_x + K_S \sigma)^2} - \frac{K_S (1 + l_x)}{(1 + l_x + K_S \sigma)^2} \right) = 0 \tag{4.26}$$

$$\frac{\partial f_{hyb,gain}}{\partial K_S} = \frac{k_t \beta_m}{d_P d_m} \left(\frac{\sigma l_x}{(l_x + K_S \sigma)^2} - \frac{\sigma (1 + l_x)}{(1 + l_x + K_S \sigma)^2} \right) = 0 \tag{4.27}$$

Solving equation 4.26 and 4.27 we obtain:

$$\sigma \cdot K_S = \sqrt{l_x^2 + l_x} \tag{4.28}$$

Therefore, to achieve the maximal signal gain, the choice of intermediate depends on how leaky the original system is. At the same time, the activating coefficient K_S and the turnover rate of the intermediate σ should align with the leak l_x based on equation 4.28. Combine the findings from equation 4.23 and equation 4.28, we learned that although σ and K_S would contribute to decreasing leak, they are also the key parameters that govern the optimization of induction strength. Therefore, in order to reduce leakage of any inducible operon, the ideal intermediate has low leak coefficient l_s and the construct should use a weak promoter (low θ) to transcribe the promoter gene. We know that l_S of STAR is usually very low and also defined by the STAR pair. However, θ is a independent of

the signal gain of this hybrid construct and it is very tunable. The promoter that sits upstream of target and GFP is constitutive and can be easily substituted to minimize leak. This architecture has offers two advantages: first, it offers a tuning point of the system. The constitutive promoter upstream of the target gene now controls the transcription the reporter gene, and it can be easily swapped out for promoters with alternative strengths. Secondly, it groups the transcription factor and the inducible promoter into a module that's isolated from the impact of downstream tuning. Theoretically, his intermediate could be any transcriptional RNA activators.

To test our theory, we next took the same xylose operon (Figure 4.2A) and implemented this design in *E.coli*. We built several versions of a Xyl-STAR hybrid operon based on the design in Figure 4.1A, with different constitutive Anderson promoters of various strengths driving the transcription of target and GFP. As shown in Figure 4.2C, the TF module (The TF xylR and the xylose inducible promoter P_{xyl}) are kept intact on the ColE1 backbone and controls the expression of STARs. The reporter gene GFP was configured in a lower copy P15A backbone, and regulated by STAR's counter part Target. Each version of the construct has a version of constitutive Anderson promoter (highlighted in orange). Based on equation 4.23, as the strength of this promoter goes down is essentially reducing the parameter θ. As we performed an *in vivo* culture induction, we found the theoretical prediction to be accurate: As the promoter strength goes down, the operon's leak reduces while the dynamic range stayed relatively consistent.

4.2.3 STAR amplifies dynamic range

We tested variation of our designs to verify whether tuning down the constitutive promoter would reduce leak while preserving the original functionality of the inducible

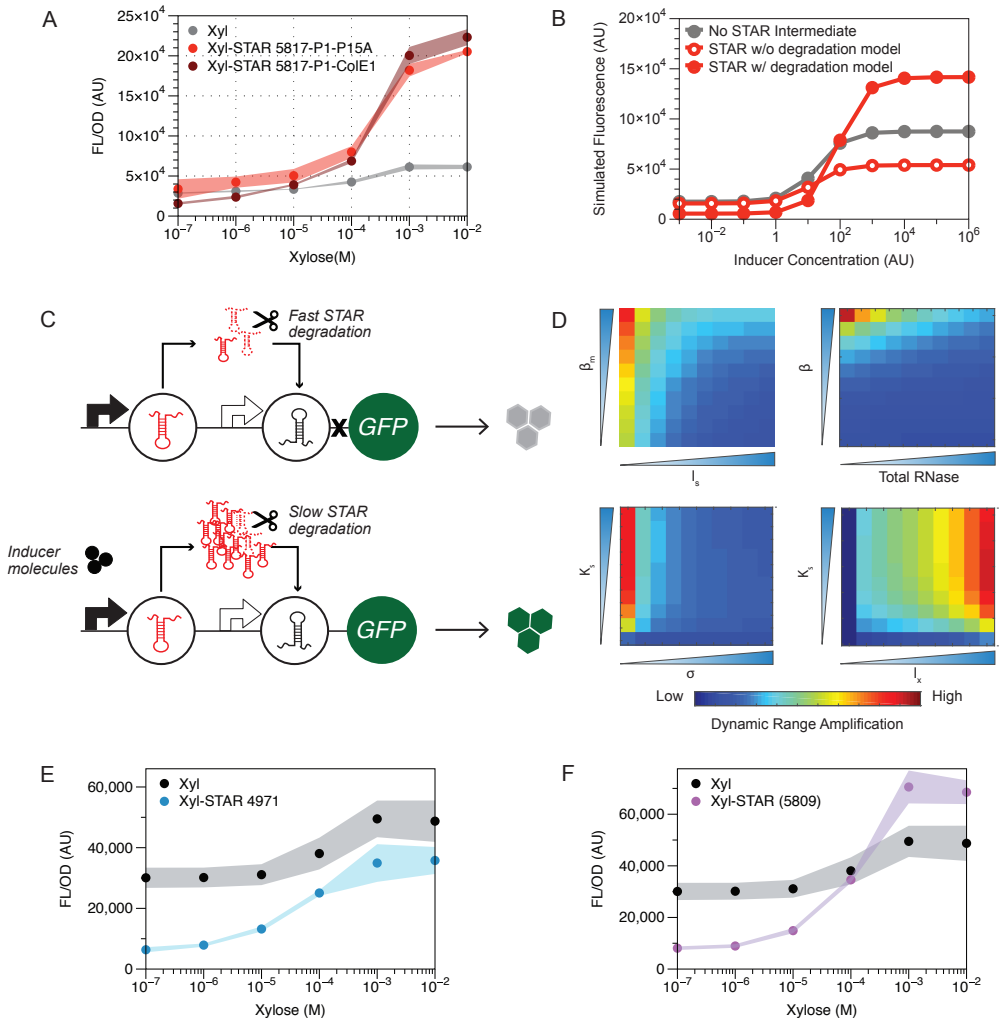


Figure 4.3: STARs amplify dynamic range of inducible systems (A) Dynamic range amplification was observed when configured with high strength constitutive promoter (orange promoter in 4.2C). Dynamic range was amplified from 2 fold to 6 folds when reporter gene was expressed via a medium copy plasmid P15A (red),further amplified to 10 folds when expressed via a high copy plasmid ColE1 (maroon). (B) Simulation results from Appendix C note 4. Simulated dynamic range decreased When the hybrid operon was modeled with constant degradation rate (hollow circles). Simulated dynamic range showed improvement with the same set of parameters when modeled with constant RNase pool. (C) the schematics of STAR mediated dynamic range amplification (D) Parameters of systems impact dynamic range amplification effects. (E) Dynamic range amplification induced by STAR 4971. (F) Dynamic range amplification induced by STAR 5809.

operon. We then titrated up the promoters to to see if it would do the opposite. However we discovered something unexpected and very interesting. We found that when we used the consensus promoter to drive GFP expression, the STAR hybrid *xyl* operon showed a significant dynamic range improvement from the original *xyl* operon, from 2 folds to about 6 folds (Figure 4.3A, red). When we increased the promoting strength even more by moving the Target-GFP to the high copy plasmid ColE1, the dynamic range was even more drastically improved to 14 folds. We then investigated which parameters were governing this effect and the molecular mechanistic behind this observation with a theoretical approach. The fold dynamic range $f_{hyb,DR}$:

$$f_{hyb,DR} = \frac{G_{hyb,on}}{G_{hyb,off}} \tag{4.29}$$

$$= \frac{\left(\frac{1+l_x}{\sigma K_S+1+l_x} + l_S\right)}{\left(\frac{l_x}{\sigma K_S+l_x} + l_S\right)} \tag{4.30}$$

To demonstrate the hybrid system can amplify dynamic range, we take the ratio of $f_{hyb,DR}$ 4.30 and $f_{x,DR}$ 4.21 and set the ratio to be greater than 1:

$$\frac{f_{hyb,DR}}{f_{x,DR}} = \frac{(\sigma K_S l_S + l_S + l_x l_S + 1 + l_x)(\sigma K_S + l_x) l_x}{(\sigma K_S + 1 + l_x)(l_x + l_S \sigma K_S + l_x l_S)(l_+ 1)} > 1 \tag{4.31}$$

Through algebraic manipulation, we obtain:

$$-l_x^2 - l_x - l_S(K_S\sigma)^2 - l_S K_S \sigma - l_x^2 l_S - l_x l_S - 2 l_x l_S K_S \sigma \not> 0 \tag{4.32}$$

Equation 4.32 suggests the dynamic range should theoretically never be improved when fusing STARs to an inducible system under current assumptions. This mismatch between theory and experimental result suggested that one or more of the current assumptions that we built our model upon is incorrect. Under current assumption, all STAR molecules degrade at the same rate d_S at all time, which might not hold true if RNA degrading machineries becomes limited. Because STARs are small RNA molecules,

and can be rapidly transcribed, the quantity of STAR could saturate the free RNases and cause a slow down in RNA degradation.

We redefined the RNA turnover constants in the equations:

$$\sigma_{on} = \frac{d_{S,on}}{\beta_S} \tag{4.33}$$

$$\sigma_{off} = \frac{d_{S,off}}{\beta_S} \tag{4.34}$$

Assuming the mRNA degradation rate is not impacted for simplicity, we substituted equations 4.33 and 4.34 into equation 4.31, we obtain:

$$\frac{f_{hyb,DR}}{f_{x,DR}} = \frac{\frac{l_x+1}{K_S\sigma_{on}+l_x+1} + l_S}{\frac{l_x}{K_S\sigma_{off}+l_x} + l_S} \cdot \frac{l_x}{l_x+1} \tag{4.35}$$

We further simplified equation 4.35 with the assumption that $l_S = 0$, yielding:

$$\frac{f_{hyb,DR}}{f_{x,DR}} = \frac{K_S\sigma_{off} + l_x}{K_S\sigma_{on} + l_x + 1} \tag{4.36}$$

In order for dynamic range to improve, equation 4.36 need to be > 1. Through algebraic manipulation, we arrive at

$$K_S(\sigma_{off} - \sigma_{on}) > 1 \tag{4.37}$$

According to equation 4.37, in order for dynamic range to increase, STAR molecules have to have a faster turn over rate at off state and a slower turn over rate σ at on state. Note that the turn of rate is $\sigma = d_s/\beta$, which is the ratio of degradation rate and the transcription rate of STAR molecules. Since β is defined as the maximum transcription rate of the inducible promoter at on state, this parameter is not subjected to change by definition. The only plausible scenario where equation 4.37 holds true is a RNA degradation slow down due to RNA overproduction.

Next we simulated equation 4.1 to equation 4.3 with a constant RNase pool that dictates the degradation rate (Appendix C4.3). The simulation suggested that, with the

same set of parameters, if degradation is independent of the amount of free RNase, the dynamic range (Appendix C4.2) of the operon goes down. On the other hand, if RNA degradation is modeled to be governed by a constant pool of total RNase (Appendix C 4.2), we observe a significant dynamic range improvement (Figure 4.3B).

This mechanistic is easy to translate into intuition. As showed in figure 4.3C, if the inducible promoter is leaky, some small amount of STARs are transcribed at the off state when inducer is absent. These STARs are quickly degraded so that it does not reach the Target's activating threshold, therefore, the GFP gene is not transcribed therefore the leaked signal is low. While at the on state with the presence of the inducer molecules, STARs are being rapidly transcribed. As it reaches a certain threshold, the large quantity of STAR molecules would saturate the free RNases and causes a slowdown in RNA degradation. This slowdown does not only stabilize STAR, it also stabilizes the mRNA. This RNA stabilization effect acauses the a signal amplification at on state.

If the over production of STAR molecules lead to dynamic range amplification, why is this phenomena only observed when the Target-GFP is transcribed by high strength promoter (Figure 4.2C)? What other criteria govern this amplification? To answer these questions, we simulated the dynamic range amplification with ranges of key parameters and presented them in heat maps (Figure 4.3D). The first heap map (upper left) was plotted with the constitutive promoter strength β_m and STAR's leak coefficient l_S. This plot showed that as the dynamic range amplification decreases as β_m goes down, and as the leak coefficient of STAR l_S goes up. This is because when the constitutive promoter is weak, it defines a low ceiling for STAR activation. STARs are being overproduced and stabilized at on state regardless the promoter strength of reporter GFP, but this amplified signal cannot be transmitted to the reporter level if the signal is being capped by a weak

promoter. At the same time, the second plot (upper right) showed the impact of the RNase pool size and the strength of the inducible promoter β, which defines the amount of STARs produced at on state. Because the key of dynamic range amplification is mainly caused by the saturation of RNA degradation machineries, it requires a strong β and a relatively small RNase pool. This suggests that if an operon has weak promoting strength, although STARs might still help tuning the signal gain of the promoter, it is unlikely that the dynamic range would be amplified. The 3rd plot (lower left) indicates that the property of the STAR intermediate is also crucial for amplification. Specifically, the RNA turn over rate σ and the activation coefficient K_s. It showed that smaller σ, which translates into slow sRNA degradation rate, and larger K_s favor dynamic range amplification. A relatively slow degradation can effectively cause slowdown in RNA degradation at on state, and a larger K_s could threshold the leak better at off state. The 4th figure (lower right), also looked at the leak from the original operon l_x. This suggests that this method would result in a more significant improvement if the original operon is leaky. If the original operon has high dynamic range, the improvement would likely to be less apparent.

Based on our understanding of the dynamic range amplification effect, we hypothesized that all three pairs of STARs woulwd amplify dynamic range of the xylose operon. We created two more different versions of xylose-STAR hybrid operons and and tested them for comparison. As shown in Figure 4.3E and 4.3F , both operons improved the dynamic range of the original xylose operon. STAR 4971 drastically lowered the leak; STAR 5809 decreased the leak and improved the on state. This showed that although differ by transfer function, the improvement of dynamic range is universal across all 3 pairs of STAR.

4.2.4 STAR sequesters also impact transfer function and dynamic range

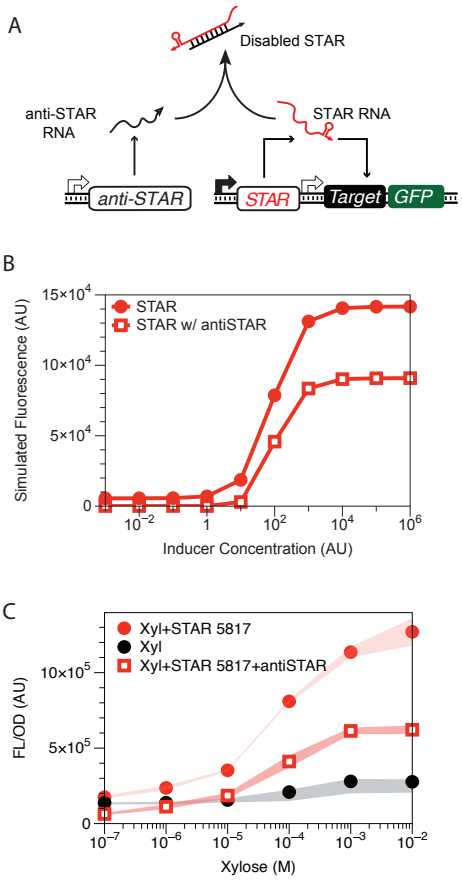


Figure 4.4: STAR sequesters impact transfer functions (A) Schematics of the hybird operon configured with STAR sequester. (B) Simulated transfer function shift when hybrid operon configured with STAR sequester. The on state have a larger signal deduction than the off state due to RNA degradation slowdown (C) Experimentally collected transfer functions. Comparing xylose operon (grey, 2 fold),xylose STAR hybrid (red solid dots 10 fold) and hybrid operon configured with sequester (red hollow circles 14 fold). The experimental result agreed with simulation that the ON state signal suffered a larger deduction than the OFF state.

STARs were all designed to have a similar structures: a short piece of RNA that folds into a linear region following an RNA stem. The linear region of the RNA is responsible for looking for it's Target component which has the complementary sequence. This

region can also be targeted by the STAR sequester, which is a shorter piece of sRNA that is the complementary of the STAR's linear sequence. Once the sequester binds to a STAR molecule, STAR is no longer able to bind to Target, therefore is deactivated. Sequesters are known to impact transfer functions, mostly in lateral directions. Here we utilized it to reduce leak and amplify dynamic range.

As we have shown in figure 2, if we use a weak promoter to drive the transcription of GFP in the STAR-hybrid configuration, we can tune down the leak of the operon but it does not offer any dynamic range amplification.; If we use a strong promoter to drive GFP, the dynamic range of the system is improved but the leak would likely to increase (Figure 4.3A). We designed the a new configuration that involves STAR sequesters to tune down and leak as well as further improves the dynamic range amplification of the STAR hybrid operon (Figure 4.4A). This configuration adds a STAR sequester (anti-STAR) in addition to the same construct from figure 3A. The anti-STAR is constitutively expressed. An intuitive prediction was the design was to engineer an equal signal reduction at on and off level. Since the dynamic range of the system is defined as the product of on state signal and the off state signal (on state/off state), this would result in lower leak as well as amplification of dynamic range.

4.However, the sequesters also operate in the environment where RNA degradation slows down at the on state. We simulated the transfer function of the configuration with a constant RNase pool (Appendix C5). As a result, although the leak is indeed reduced at the off state, the on state suffers a larger reduction. This is also caused by RNA degradation slowdown at on state. This effect does not only stabilizes STARs, it also stabilizes the sequesters. The sequesters are more effective when the signal is high, therefore causes an uneven signal reduction. This was also observed in experimental results as shown in figure 4.4C. Although the signal reduction at on and off states are

uneven, this design still lead to a further dynamic range improvement from 8 folds to 14 folds.

4.2.5 Broad applications and limitations

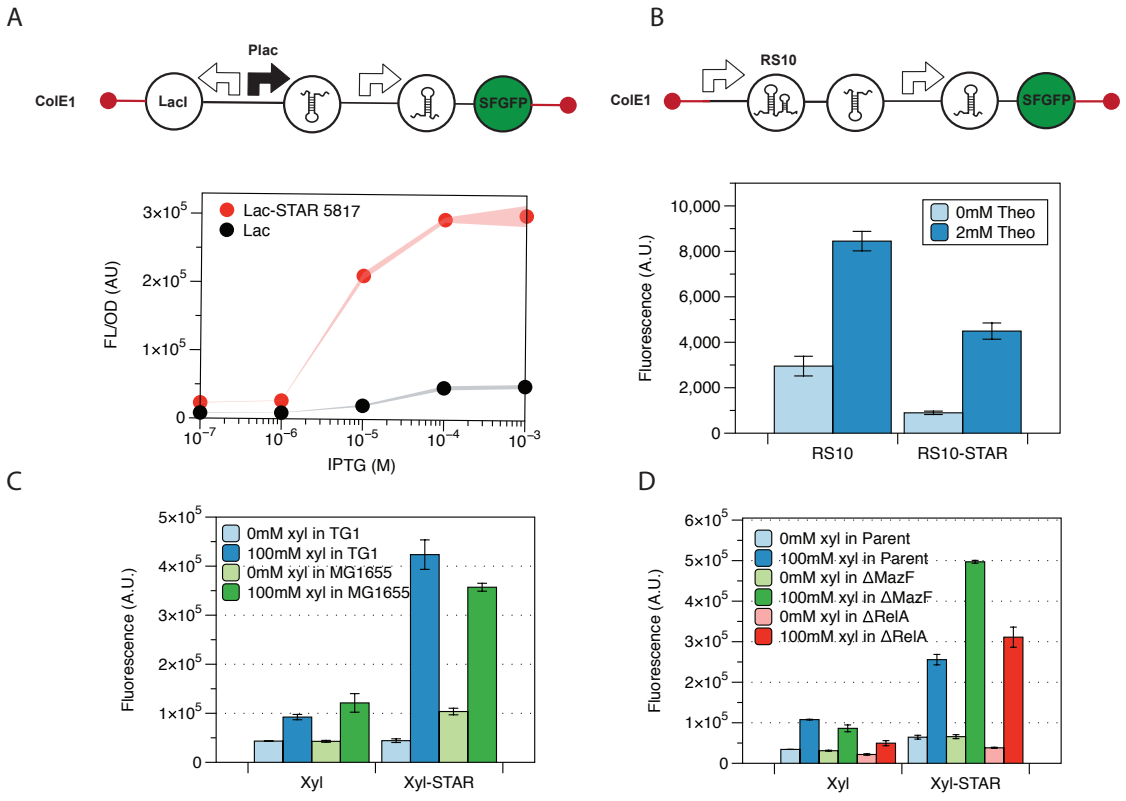


Figure 4.5: Broad application and limitation of STAR as dynamic range amplifier. (A) STAR improves a house-assembled Lac operon from 4 folds to 12 folds. (B) STAR improves a theophylline inducible transcriptional riboswitch from 2.7 fold to 5 folds. (C) The Xyl-STAR hybrid operon's dynamic range amplification effect in TG1 was not observed in MG1655. (D) STAR did not improve the dynamic range of xylose operon in the Keio parent strain. Knocking out either MazF or RelA recover the amplification effect.

There are many different versions of transcriptionally inducible activators, each with distinct mechanistic properties. However they can all be simplified with a hill function of

the inducer like we did in equation 4.16. We so far have built two versions of STAR hybrid operons with *Lux* operon and *Xyl* operon. Both of these operons are transcriptional factor based and operate with a positive activation mechanism. The TF (*XylR* or *LuxR*) binds to the inducer to form an activating complex, that sequentially activates transcription leading by the inducible promoters (P_{xyl} or P_{lux}). On the other hand, *Lac* operon operates in a different fashion. the TF *LacI* represses the transcription lead by P_{lac},the inducer binds to the repressor *LacI* and sequestrates it away to activate the transcription. Although these two types of activators function differently, they can be modeled with identical hill functions (Appendix C3). As we stated previously, one advantage of the STAR hybrid operon configuration is that the original operon is insulated and modularized, therefore the application of STAR to the operon is circuit context independent and extendable. Therefore should have a broad application to other inducible systems.

We tested this broad application potential with two different systems of distinct mechanisms. First we house-assembled an un-optimized *Lac* operon for the purpose of representing an under-performing repressor-sequestrating activator. Then we used it to built a *Lac* STAR 5817 hybrid operon on ColE1 (Figure 4.5A). The resulting transfer function improved the dynamic range from 4 folds to 12 folds. Finally, we built a riboswitch STAR hybrid As shown in figure 4.5B. This riboswitch RS10 is inducible by theophylline[120]. The RNA folds into a transcriptional terminator structure without inducer to prevent transcription of its downstream gene. When theophylline is present, theophylline binds into the RNA structure to opens up the termination hairpin, which activates transcription. We placed STAR downstream of RS10 and insulated them with a piece of ribozyme to prevent structure interference[46]. In this experiment we observed a leak reduction as well as a dynamic range improvement from 2.7 fold to 5 fold.

We learned from the figure 4.3 that STARs can serve as a dynamic range amplifier

if the overproduction of STAR at ON state causes an RNA degradation slowdown. This suggests that the amplification effect wouldn't be observed in all cellular environments. For instance, if the total RNase pool becomes much more abundant, the on state might not trigger the RNA degradation slowdown for amplification. We hypothesized that different bacteria strains would have different cellular environments therefore the RNase pool size might be different. We then repeated the same experiment in Figure 4.3A in strain MG1655 to access the dynamic range of the hybrid operon in a new strain. As a result, the dynamic range is much smaller in MG1655, however, it did not match the behavior profile of a larger RNase pool. As shown in figure 4.5C, the OFF state signal of the STAR hybrid operon in MG1655(green bars) is much higher than its OFF state signal in TG1 (blues). This suggests that the degradation rate is slower in MG1655 at OFF state. At the same time, the ON state signal in MG1655 is lower than it is in TG1. This suggests a faster RNA degradation at ON state. This observation, therefore, cannot be explained by the simple RNA degradation machinery pool size.

MG1655 strain is one of the most commonly used *E.coli* K12 isolate descendant[10], whereas TG1 is a derivative of a K12 and B strain hybrid HB101[56]. Although they share a large amount of genomic DNA, they have very distinct phenotypes. TG1 is known to be fast growing, which makes the strain popular in synthetic biology laboratories. However we still don't have the full picture of each strain's cellular environment. We hypothesized that this uneven RNA degradation rate we observed was caused by some stress response genes that are known to exist in MG1655 strains. We tested two stress response genes in MG1655 that might impact this amplification effect: *MazF* and *RelA*. These two genes are both stringent response genes located in the same locus and are activated in response to amino aicd starvation. *MazF* is an Endoribonuclease toxin, it cleaves single stranded mRNA on the $5'$ end of ACA sequences to inhibit translation when activated[132]. *RelA* is a (p)ppGpp synthetase that down regulates translation

through interaction with ribosomes[125]. These stringent responses are not accounted in our previous models in figure 4.3, they could be triggered by high expression of synthetic circuits. The resulting down-regulation in mRNA translation could negate the dynamic range amplification caused by STARs at the transcriptional level. To test this theory, we repeated the same experiment in figure 4.4A (maroon) and Figure 4.5C in three strains in the Keio collection: the parents strain that's derived from MG1655 (BW25113,[28]), the strain with a single *MazF* knockout (JW2753-1,[4]) and the strain with a single *RelA* knockout (JW2755-3,[4]). As a result, we found that the experiment performed in the parent strain displayed a very similar profile with MG1655: the hybrid operon has a higher OFF level expression than the *xyl* operon and the dynamic range stayed static. On the other hand, two strains with *MazF* and *RelA* knockouts demonstrated dynamic range amplification induced by STAR hyrbid operon. This observation confirmed our hypothesis on the unsatisfying amplification result in MG1655. Because we don't have the exact genomic profile of TG1, we can only speculate that its stringent response *RelA* operon is either absent or disrupted.

4.3 Conclusion

In conclusion, we uncovered the design principle of transfer function manipulation actuated by STARs with the combination of theory, simulation and experiments. Our investigation was sprouted out of an experimental observation in our previous work on sTARs[22]. Where we fused STAR as a intermediate between an AHL inducible *Lux* operon and reporter GFP to construct a STAR hybrid operon, and we observed transfer function shift across our variants. We first use effective models to identify and estimated the contributing parameters to this shift based on our previous modeling method [48, 46]. The resulting parameters simulated response curves that highly resembles our previous experimental observations. Next, we presented that a major advantage of this circuit configuration is a supplemented tunning knob at the transactional level that is effective for reducing signal leakage at OFF stat. At the same time, We also discovered the the dynamic range of the original operon was amplified when this tuning knob was turned up, which was not predicted by our original model. We investigated into this observation again and found the underlying causes of this phenomenon. We found that this amplification was caused by a RNA degradation slowdown caused my overproduction of STARs at on state. Furthermore, we added in another layer of STAR sequester and showed that it also impacts transfer function and potentially amplifies dynamic range. In addition, we applied the same configuration to several other inducible systems and showed the dynamic amplification is generic and extendable. Finally, we discussed the important role of cellular environment in the amplification effect and demonstrated various experimental results of the same circuit in different *E.coli* strains. We showed that the stress response network in some *E.coli* strains would interfere with the synthetic circuit and negate its dynamic range amplification effect.

This work is an investigation into dynamic range of inducible systems with the com-

bination of theory, simulation and experiments. Several aspects of this work are significant. First of all, this work is a new application of the effective model and parameterization method[48] on understanding the effect of regulatory RNA on protein mediated systems. We successfully parameterized STAR for the first time, the insights provided by these parameters helps us understand an transfer function shift we previously observed but didn't comprehend. Second, we presented that this configuration adds an tuning knob to inducible systems otherwise have static transfer functions. This could widen the application of numerous inducible systems in complex multilayer synthetic gene circuits. Third, we discovered the dynamic range amplification properties of this configuration, and we identified the underlying causes of this effect.At the same time, it is the first time sequesters are used to manipulate transfer functions vertically. Finally, by presenting its extendability and limitations, we uncovered the design principle of this transfer function manipulation, which will guide future designs of dynamic range amplifiers with STARs. One of the major aspects of this design principle interfaces with the cellular environment that hosts synthetic circuits. Synthetic circuits operate in the inner environment of living cells. As part of the biological network, it's subjected to environmental changes and constrains as well. This configuration is an example of optimizing the performance of synthetic circuits with the cellular context. Overall, this work not only offers a versatile tool that manipulates transfer function and amplifies dynamic range, it also is an example of how theory and simulations combined with experiments leads to a better understanding of synthetic biology.

4.4 Material and Methods

Plasmid construction and purification

The STAR variants 4971, 5809, 5816 were from Chappell et al[22]. The xylose operon was obtained from BBa733002 in the iGEM registry. The Riboswitch RS10 sequence was from Wachsmuth at al [120]. All constructs were created using Golden Gate assembly[32] and Gibson assembly. Plasmids were purified using a Qiagen QIAfilter Plasmid Midi Kit (Catalog number: 12243) followed by isopropanol precipitation and eluted with double distilled water.

4.4.1 Model simulation

All models were simulated by solving corresponding ODEs using Matlab function ode15s over a set of discrete time steps using guessed or estimated parameters. All initial conditions for species concentrations were zero.

4.4.2 TX-TL experiments

TX-TL buffer and extract was prepared according to Garamella et al [36]. TX-TL buffer and extract tubes were thawed on ice for approximately 20 min prior to performing experiments. Separate reaction tubes were prepared with combinations of DNA representing a given network condition. Appropriate volumes of DNA, buffer, and extract were calculated using a custom spreadsheet developed by Sun et al [106]. Buffer and extract were mixed together and incubated for 20 min at 37C. DNA for the spe-

cific experiment was then added into each tube according to the previously published protocol[109, 48]. Ten L of each TX-TL reaction mixture was transferred to a 384-well plate (Nunc 142761), covered with a plate seal (Nunc 232701), and placed on a Biotek SynergyH1m plate reader. Temperature was controlled at 29C. Superfolder GFP (SFGFP) (485 nm excitation, 520 nm emission, 100 gain) or malachite green (MG) (610 nm excitation, 650 nm emission, 150 gain) fluorescence was measured every 5 min for 120 min. MG constructs were observed to immediately generate observable signal in between mixing and loading the plate reader, making quantification of early time points difficult. Therefore, the first measurements of all constructs with MG were assumed to be 0 and data collection was timed to start 5 minutes after reactions were mixed. Each reaction was run with a minimum of triplicate technical repeats, and repeated three times independently (minimum of nine total replicates). A ten L sample of TX-TL buffer, extract, and water was run together with each independent experiment as blank. All data points were then processed with blank values subtracted at each time point.

4.4.3 Parameter estimation

Parameters were estimated from each designed experiment by fitting the identifiable parameters of that experiment to measured fluorescence trajectories. The parameter estimation problem is given by:

$$min(G_{t_k}^{experimental,mean} - G_{t_k}^{simulated})^2 \tag{4.38}$$

Where $G_{t_k}^{experimental,mean}$ denotes the average value of the experimentally observed SFGFP expression at a certain time t_k. The vector P contains all of the identifiable parameters in the experiment being analyzed. $G_{t_k}^{simulated}(P)$ denotes the model simulated SFGFP expression at time t_k. To optimize the initial guesses, we first started with a set of initial

guess based on pervious work [48]. We then constructed uniform distributions (100 points) around each parameters value (15%), generating 100 sets of parameters, which led to 100 simulated dynamic trajectories for each of the 6 estimation experiments designed by sensitivity analysis procedure. We picked the set of parameters that generated the best fit to the 6 sets of experimental data collectively, and used it as a starting point for the next round. We repeated this procedure 10 times to achieve the best guess. We then started the estimation process with the best-guessed parameter set and constructed uniform distributions (1000 points) around each parameter value (15%). Each set of parameters served as a different starting point for finding optimal parameter estimates. Each optimal set was found by sequentially applying the fitting procedure to each of the designed parameters using the *Matlab* function *fmincon*, only fitting the identifiable parameters for that experiment. Fit parameters from one experiment were then used to update and replace the initial guesses before moving on to the next experiment until each of the 16 unknown parameters were fit from the 6 experiments. The result of this procedure produces 10000 sets of estimated parameters, which were then subject to the analysis described in the main text.

4.4.4 Strains, growth media and *in vivo* gene expression experiments

All experiments in the first four figures were performed in *E. coli* strain TG1. Some experiments were performed in other strains as noted in figure 5. Plasmids with different network constructs were transformed into TG1 competent cells, plated on LB + Agar plates containing 100 g/ml carbenicillin, and incubated overnight at 37°C. At least three colonies of each experimental condition were inoculated into 300 L of LB containing

carbenicillin in a 2mL 96-well block (Costar 3960), and grown for 12 hours overnight at 37°C at 1000 rpm in a Labnet Vortemp 56 bench top shaker. Four microliters of the overnight culture were added to 196 l of M9 supplemented M9 media (1X M9 minimal salts, 1 mM thiamine hydrochloride, 0.4% glycerol, 0.2% casamino acids, 2 mM MgSO4, 0.1 mM CaCl2) containing carbenicillin and grown for 5 hours at 37°C at 1000 rpm. Fluorescence (485 nm excitation, 520 nm emission) and optical density (OD, 600 nm) were then measured on a Biotek SynergyH1m plate reader. For flow cytometry measurement, the sub-culture was then diluted to 0.015 OD with M9 + carbenicillin and 100nM N-Acyl homoserine lactones (AHLs) to a total volume of 3ml and grown in culture tubes for 5 hours at 37°C at 220rpm. A total of 50,000 events were collected on a BD AccuriTM C6 plus Flow Cytometer during measurement. All experiments contained three replicates from at least three different transformed colonies, and were repeated independently three times for a minimum of nine total replicates.

CHAPTER 5

CONCLUSION AND PERSPECTIVES

5.1 Conclusion

In this work, not only have we established the foundation of mathematical and theoretical models for RNA circuitry, we have also made significant advances in RNA synthetic biology with the guidance provided by these models. First, we presented the first validated computational model of synthetic RNA transcriptional genetic circuitry. This work helps lay the foundation for CAD tools that can incorporate RNA regulators into synthetic circuitry designs, and opens up the possibility for more advanced RNA circuits to be engineered. Then we demonstrated the utility of using effective models in combination with TX-TL experiments to systematically and rapidly prototype RNA network designs. Using this design and prototyping method, we were able to create the first functional sRNA mediated transcriptional NAR network that functions both in TX-TL and *in vivo*. During this process, we established a streamlined and quantitative model guided design and TX-TL based prototyping methodology for in-cell RNA network engineering. We anticipate that this design-build-prototype-test network building process will minimize cloning effort and speed up network development cycles as the field moves towards more sophisticated synthetic network engineering. Finally, we were able to use the effective model and the sensitivity analysis based parameterization method to uncover the design principle of transfer function manipulation using STARs. The interdisciplinary method that theory, simulation and experiments unveiled the impact of context of cellular environment on synthetic circuits.

5.2 Future perspectives

The power of regulatory RNA has been slowly revealed by synthetic biologists in the past decade. Aside from Lucks lab's progress in engineering transcriptional antisense RNAs, numerous other RNA regulators have been developed and improved. *Green et al.*[40] developed a large library of translational regulatory RNA called toehold switches. It has protein-like dynamic range and have the potential to sense any mRNA. The Pierce lab [43] also developed small conditional RNA (scRNA) that can be used to perform signal transduction between diverse programmable inputs and outputs. It is the beginning of an exciting era for RNA synthetic biology. There is a tool box of powerful, programale and versatile regulatory RNAs available for RNA circuit engineering. The biggest roadblock we face today is a reliable computer aided design (CAD) tool that's capable of predicting the fast dynamics of RNA circuitry.

This work took a first step towards creating a reliable CAD tool. However, we have long way left towards this goal. First of all, we need to build models that accurately predicts the behavior of RNA circuitry *in vivo*. The current model is based on circuit dynamics in TX-TL. although TX-TL has a promising application in diagnostics, it is still an over-simplified mock cellular environment. Many elements including extrinsic noise, cell fate, stress, genomic network interference are not included in the current model. In order to predict RNA dynamic accurately *in vivo*, a stochastic model is required to incorporate the stochastic noise of gene expression in living cells. At the same time, experiments that track circuit dynamic of single cells are also essential for us to understand and eventually model some unpredictable cellular behaviors. Secondly, we need to learn how to model these circuits with resource constraints. Our current model in TX-TL environment was built under the assumptions of constant rates and unlimited resources. It is a valid assumption when TX-TL reactions run for a short period of time

with a small enough of DNA load. To expand this window, we need to incorporate energy depletion and the limitation of gene expression machineries. Finally, new models that consider the interface between synthetic circuits and cellular networks are essential for accurate *in vivo* dynamic predictions. We have explored some of these properties in chapter 4, but a full profile of stress response in bacteria is yet determined or incorporated into models.

APPENDIX A

APPENDIX OF CHAPTER 2: GENERATING EFFECTIVE MODELS AND PARAMETERS FOR RNA GENETIC CIRCUITS

To model the 3 level sRNA transcriptional cascade, we started with the simplest equations possible and added complexity when required to capture features of the sRNA transcriptional repressor system. To start, we used four ordinary differential equations to describe the time varying concentrations of four species (sRNAs R_2 and R_1, messenger RNA of protein, M, and protein, G) as a function of time (Equations A.1-A.4):

$$\frac{dR_2}{dt} = \beta_2 - d_2 \cdot R_2 \tag{A.1}$$

$$\frac{dR_1}{dt} = \beta_1 \cdot (1 - \frac{R_2}{K_2 + R_2}) - d_1 \cdot R_1 \tag{A.2}$$

$$\frac{dM}{dt} = \beta_m \cdot (1 - \frac{R_1}{K_1 + R_1}) - d_m \cdot M \tag{A.3}$$

$$\frac{dG}{dt} = k_t \cdot M - d_g \cdot G \tag{A.4}$$

where β are transcription rates, d are degradation rates, and kt is the translation rate of the relevant species. In these equations, the transcriptional repression mechanism was described by a first order Hill function, with parameters K describing the sRNA concentration range required for repression (Figure A.1A).

Next we incorporated crosstalk into the model. Despite our best efforts to develop completely orthogonal sRNA repressor/attenuator pairs, a small amount of crosstalk always exists[110].To incorporate this feature, we modeled crosstalk using a constrained fussy logic formulation, treating the bottom level of the cascade as an OR gate module with two signaling species accounting for the contributions of cognate and cross-talk interactions[71] (Figure A.1B). We only considered cross talk between R_2 and A_1 since R_2 was on the level-3 (L3) DNA that was 28 times as concentrated as A1 on the level-1

(L1) DNA in TX-TL reactions. Accordingly we assumed the crosstalk between R_1 and A_2 to be negligible because they are expressed from the same DNA. The resulting ODE set is shown in equations A.5-A-5.8

$$\frac{dR_2}{dt} = \beta_2 - d_2 \cdot R_2 \tag{A.5}$$

$$\frac{dR_1}{dt} = \beta_1 \cdot (1 - \frac{R_2}{K_2 + R_2}) - d_1 \cdot R_1 \tag{A.6}$$

$$\frac{dM}{dt} = \beta_m \cdot \Big[1 - (\frac{R_1}{K_1 + R_1} + \frac{R_2}{\frac{K_1}{p_c} + R_2} - \frac{R_1}{K_1 + R_1} \cdot \frac{R_2}{\frac{K_1}{p_c} + R_2})\Big] - d_m \cdot M \tag{A.7}$$

$$\frac{dG}{dt} = k_t \cdot M - d_g \cdot G \tag{A.8}$$

where p_c is the percentage crosstalk M would experience from R_2, which ranges from 0 to 1.

An intriguing feature of the sRNA transcriptional repression mechanism is that attenuators can be arranged in tandem to achieve stronger repression of the target gene[108]. In addition, it was found that the repression due to tandem attenuators can be accurately described by multiplying the repression transfer curves of individual attenuators[64]. To incorporate tandem attenuators in our model, we therefore raised the repressive Hill functions for the single attenuators on each level to the power of tandem attenuator number on that level, N (Figure A.1C), resulting in equations A.9-A.12.

$$\frac{dR_2}{dt} = \beta_2 - d_2 \cdot R_2 \tag{A.9}$$

$$\frac{dR_1}{dt} = \beta_1 \cdot (1 - \frac{R_2}{K_2 + R_2})^{N_2} - d_1 \cdot R_1 \tag{A.10}$$

$$\frac{dM}{dt} = \beta_m \cdot \Big[1 - (\frac{R_1}{K_1 + R_1} + \frac{R_2}{\frac{K_1}{p_c} + R_2} - \frac{R_1}{K_1 + R_1} \cdot \frac{R_2}{\frac{K_1}{p_c} + R_2})\Big]^{N_1} - d_m \cdot M \tag{A.11}$$

$$\frac{dG}{dt} = k_t \cdot M - d_g \cdot G \tag{A.12}$$

Here N_1 is the number of tandem attenuators in L1 and N_2 is the number of tandem attenuators in L2.

Next we incorporated the effect of auto-termination. Previously it was found that some aspect of the attenuation mechanism causes repression of the downstream gene even in the absence of any antisense sRNA[64]. While the mechanistic details of this effect are not known, this can be effectively modeled by introducing the parameter Pt, which is the probability that an attenuator auto-terminates itself. This introduces a factor of $(1 - P_t)$ for each attenuator in the cascade (Figure A.1D), resulting in equations A.13-A.16:

$$\frac{dR_2}{dt} = \beta_2 - d_2 \cdot R_2 \tag{A.13}$$

$$\frac{dR_1}{dt} = \beta_1 \cdot \left[(1 - \frac{R_2}{K_2 + R_2}) \cdot (1 - P_t)\right]^{N_2} - d_1 \cdot R_1 \tag{A.14}$$

$$\frac{dM}{dt} = \beta_m \cdot \left\{\left[1 - (\frac{R_1}{K_1 + R_1} + \frac{R_2}{\frac{K_1}{p_c} + R_2} - \frac{R_1}{K_1 + R_1} \cdot \frac{R_2}{\frac{K_1}{p_c} + R_2})\right] \cdot (1 - P_t)\right\}^{N_1} - d_m \cdot M \tag{A.15}$$

$$\frac{dG}{dt} = k_t \cdot M - d_g \cdot G \tag{A.16}$$

Next, we adjusted these equations to more accurately reflect the translation and protein degradation systems within TX-TL reactions. In particular, the TX-TL system we used has very low levels of proteases, effectively removing protein degradation5. In addition, we modeled translation as a two-step event, with a ribosome binding and initiation rate k_I, and a translation elongation rate k_E, based on a previous model of prokaryotic translation6. At the same time, we factored in the maturation rate of the green fluorescence protein, α, to account for delays in observing SFGFP fluorescence after it has

been synthesized. This resulted in equations A17-A22:

$$\frac{dR_2}{dt} = \beta_2 - d_2 \cdot R_2 \tag{A.17}$$

$$\frac{dR_1}{dt} = \beta_1 \cdot \left[(1 - \frac{R_2}{K_2 + R_2}) \cdot (1 - P_t)\right]^{N_2} - d_1 \cdot R_1 \tag{A.18}$$

$$\frac{dM}{dt} = \beta_m \cdot \left\{\left[1 - (\frac{R_1}{K_1 + R_1} + \frac{R_2}{\frac{K_1}{p_c} + R_2} - \frac{R_1}{K_1 + R_1} \cdot \frac{R_2}{\frac{K_1}{p_c} + R_2})\right] \cdot (1 - P_t)\right\}^{N_1} - d_m \cdot M - k_I \cdot M + k_E \cdot M_i \tag{A.19}$$

$$\frac{dM_i}{dt} = k_I \cdot M - k_E \cdot M_i \tag{A.20}$$

$$\frac{dG}{dt} = k_E \cdot M_i - \alpha \cdot G \tag{A.21}$$

$$\frac{dG_m}{dt} = \alpha \cdot G \tag{A.22}$$

Finally, we added equations describing the apparent maturation process of sRNA repressors from their initially synthesized states ($R^{'}$) to their functional states (R), with rates r_{m2} and r_{m1}. This resulted in the final set of equations A.23-A.30 (Figure 2.2C)

$$\frac{dR'_2}{dt} = \beta_2 - r_{m2} \cdot R'_2 - d_2 \cdot R'_2 \tag{A.23}$$

$$\frac{dR_2}{dt} = r_{m2} \cdot R'_2 - d_2 \cdot R_2 \tag{A.24}$$

$$\frac{dR'_1}{dt} = \beta_1\Big[(1 - \frac{R_2}{K_2 + R_2}) \cdot (1 - P_t)]^{N_2} - r_{m1} \cdot R'_1 - d_1 \cdot R'_1 \tag{A.25}$$

$$\frac{dR_1}{dt} = r_{m1} \cdot R'_1 - d_1 \cdot R_1 \tag{A.26}$$

$$\frac{dM}{dt} = \beta_m \cdot \left\{\left[1 - (\frac{R_1}{K_1 + R_1} + \frac{R_2}{\frac{K_1}{p_c} + R_2} - \frac{R_1}{K_1 + R_1} \cdot \frac{R_2}{\frac{K_1}{p_c} + R_2})\right] \cdot (1 - P_t)\right\}^{N_1} - d_m \cdot M - k_I \cdot M + k_E \cdot M_i \tag{A.27}$$

$$\frac{dM_i}{dt} = k_I \cdot M - k_E \cdot M_i \tag{A.28}$$

$$\frac{dG}{dt} = k_E \cdot M_i - \alpha \cdot G \tag{A.29}$$

$$\frac{dG_m}{dt} = \alpha \cdot G \tag{A.30}$$

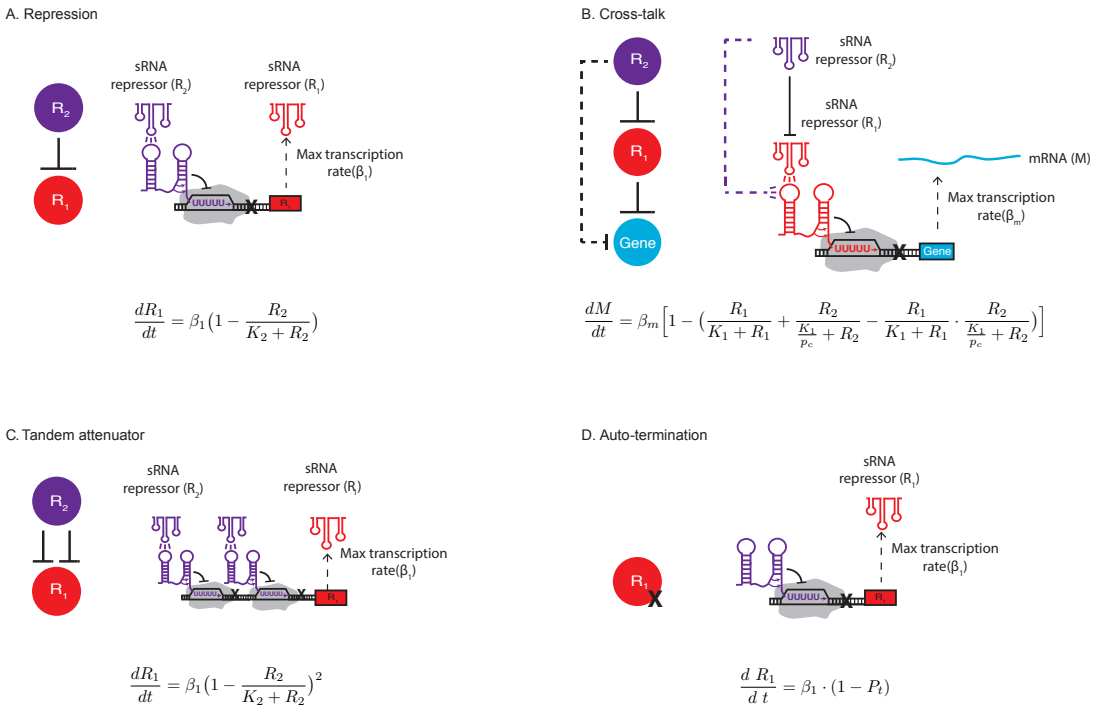


Figure A.1: Model components of the 3-level cascade. (A) Transcriptional repression governed by sRNA repressors. Presence of R_2 causes attenuator A_1 upstream of R_1 to fold into a terminator structure thus halting transcription of R1. This complex event can be described by a first order Hill function[64], characterized by a repression coefficient, K_2,which has units of concentration. K_2 defines the concentration of active R_2 needed to achieve significant repression of R_1[1]. (B) Crosstalk between repressor R_2 and attenuator A_1 can be described as an OR gate module with two signaling species[71]. Here p_c is the percentage crosstalk due to R_2 compared to the repression caused from the cognate sRNA repressor R_1. (C) Tandem attenuators can be modeled by multiplying Hill functions together[64].(D) The effect of autotermination from a single attenuator is modeled by multiplying the transcription rate by a factor of $(1 - P_t)$, where Pt is the probability of auto-termination

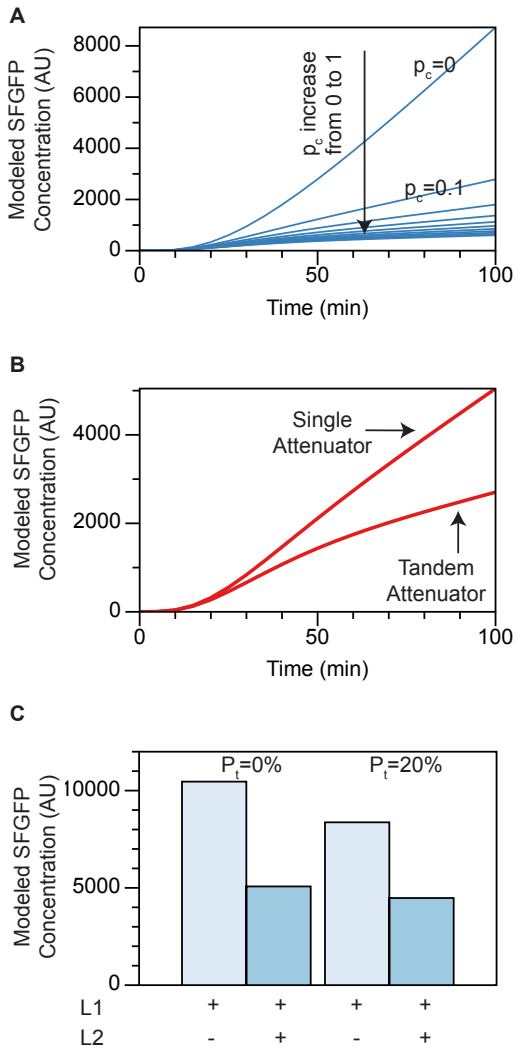


Figure A.2: Qualitative validation of model governing equations. (A) Cross-talk qualitative validation, generated by equations A.5-A.8. As cross-talk becomes stronger, the SFGFP signal is more repressed. (B) Tandem attenuator and maturation delay validation, generated by equations A.23-A.30. Both trajectories resulted from L1+L2, one with a single attenuator on L1, and one with two tandem attenuators on L1. When two attenuators are in tandem, the repression effect caused by the same repressor RNA concentration is stronger. These curves also show the characteristic bending of the trajectories at later times due to the maturation delay we factor into the model, which matches experimental characterization (Figure A.4). (C) Auto-termination validation, generated by equations A.13-A.16. Compared with no auto-termination (left two bars), 20% auto termination (right two bars) causes end point signals at 100mins to drop for both levels (L1 and L1+L2) as expected.

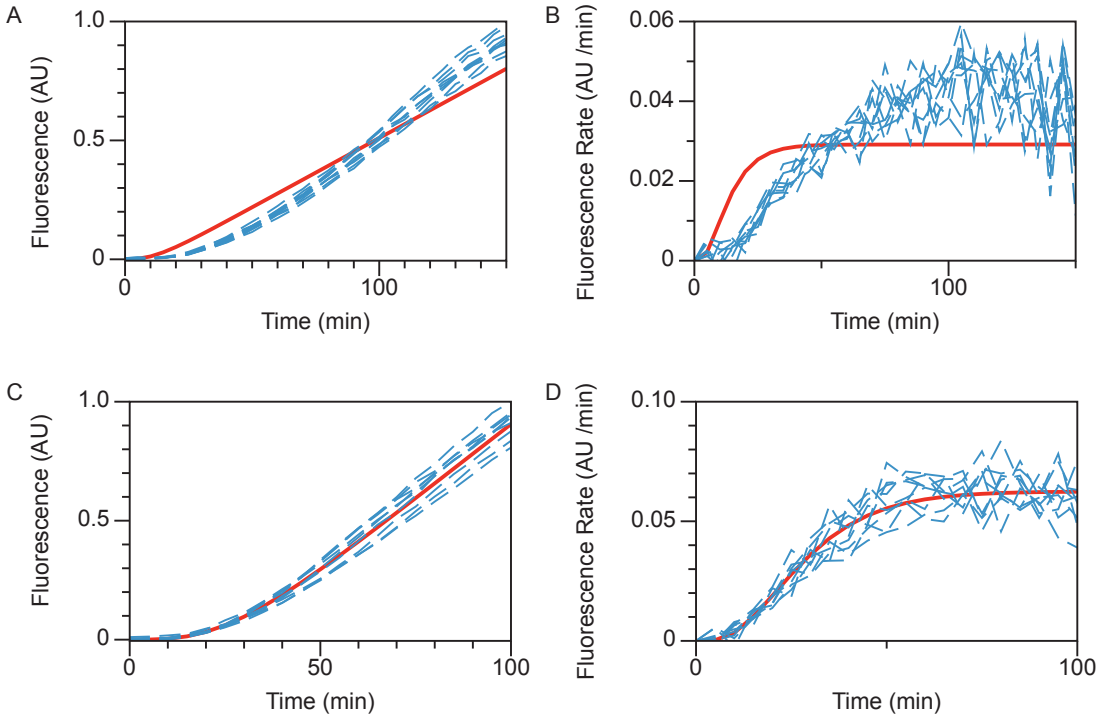


Figure A.3: Model inspired findings on pre-incubation of TX-TL reactions. (A) Best fit of the model generated L1 trajectory with equations A17-A22 (red line) and experimentally collected SFGFP trajectories (blue dash) from TX-TL reactions containing 0.5nM L1 with no pre-incubation. (B) GFP production rates derived from part A. Red line corresponds to the red model trajectory and blue dashes correspond to blue experimental trajectories. Parts A and B together show that there is a 20 minute delay in SFGPF fluorescence signal after the start of the reaction, which our model is not capable of capturing. (C) Best fit of the model generated L1 trajectory with equations A17-A22 (red line) and experimentally collected SFGFP trajectories (blue dash) from TX-TL reactions containing 0.5nM L1 with a 20 minute pre-incubation of the extract and buffer at 37°C before adding DNA at time 0. (D) GFP production rates derived from part C. Red line corresponds to the red model trajectory and blue dashes correspond to blue experimental trajectories. Parts C and D together show that a 20 minute pre-incubating of the TX-TL extract/buffer at 37°C eliminates the delay at the beginning of the reaction. Using experimental pre-incubation, our model is able to qualitatively capture the observed SFGFP trajectories.

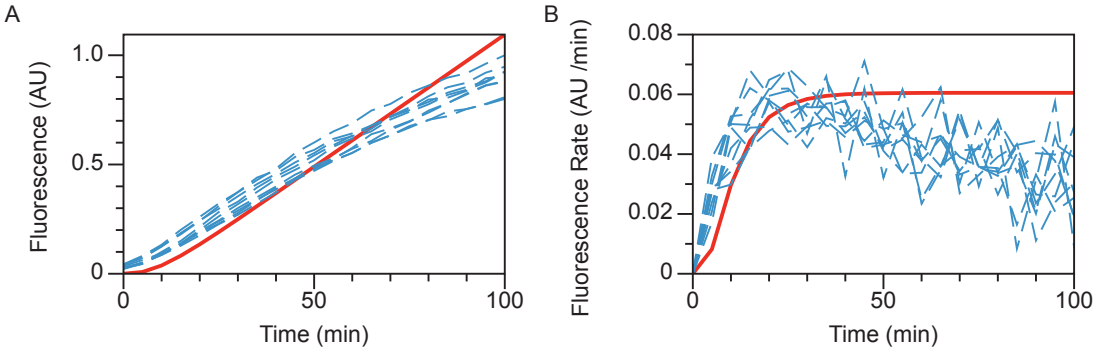


Figure A.4: Model inspired findings on sRNA maturation delay (A) Best fit of the model generated L1+L2 trajectory with equations A17-A22 (red line) and experimentally collected L1+L2 trajectories (blue dashes) with a 20 minute pre-incubation of the extract and buffer at 37°C before adding DNA at time 0. (B) GFP production rates derived from part A. Red line corresponds to the red model trajectory and blue dashes correspond to blue experimental trajectories. Parts A and part B together show that there is a drop in SFGFP production rate after the first 20 minutes in the single repression construct, which equations A17-A22 cannot capture.

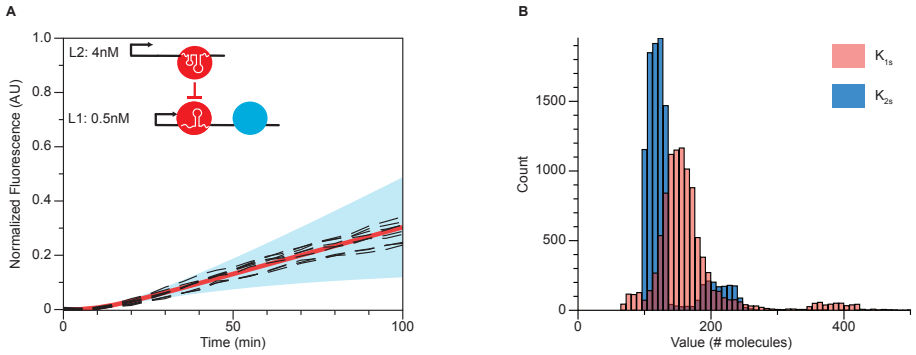


Figure A.5: Estimation of K_{1s} for cascade swap prediction. (A) The two-level parameterization experiment (black dash lines) was performed to estimate a single parameter K1s to predict trajectories for a swapped cascade circuit (Figure 2.5E,F). The mean simulated trajectory (red line) is shown within 95% confidence intervals (blue region) using the newly estimated K1s values (B) Comparison of the estimated repression coefficients between single antisense copies of R_1 (Red) and R_2 (Blue) (K_{1s} and K_{2s}, respectively). K_{2s} is lower than K_{1s} indicating that R_2 shows slightly stronger repression as has been observed previously[110, 108].

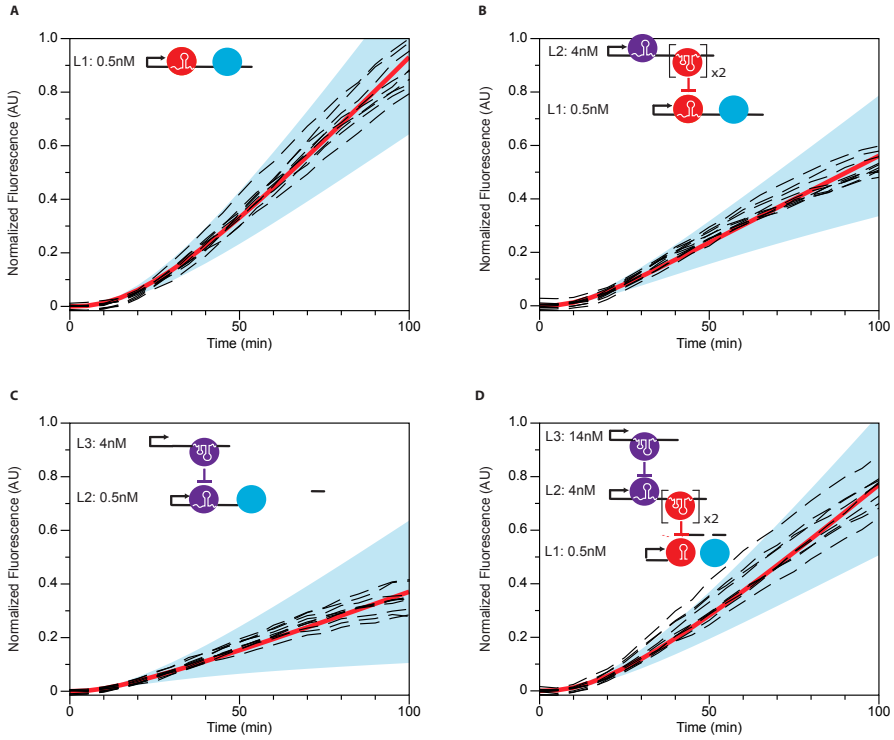


Figure A.6: Parameter estimation validation on TX-TL batch B. Comparison of experimental trajectories of SFGFP fluorescence in TX-TL experiments using batch B of extract (black dash lines) with simulated model predictions. Model simulated trajectories were generated by performing 1,000 simulations with parameters drawn from the set of 10,000 determined from the estimation procedure (see Methods). Experimental and model trajectories were normalized by the maximum observed experimental fluorescence (see Methods). The mean simulated trajectory (red line) is shown within 95% confidence intervals (blue region). The schematic of each experiment is shown in the upper left corner of each plot corresponding to the experiments in Figure 2.3.

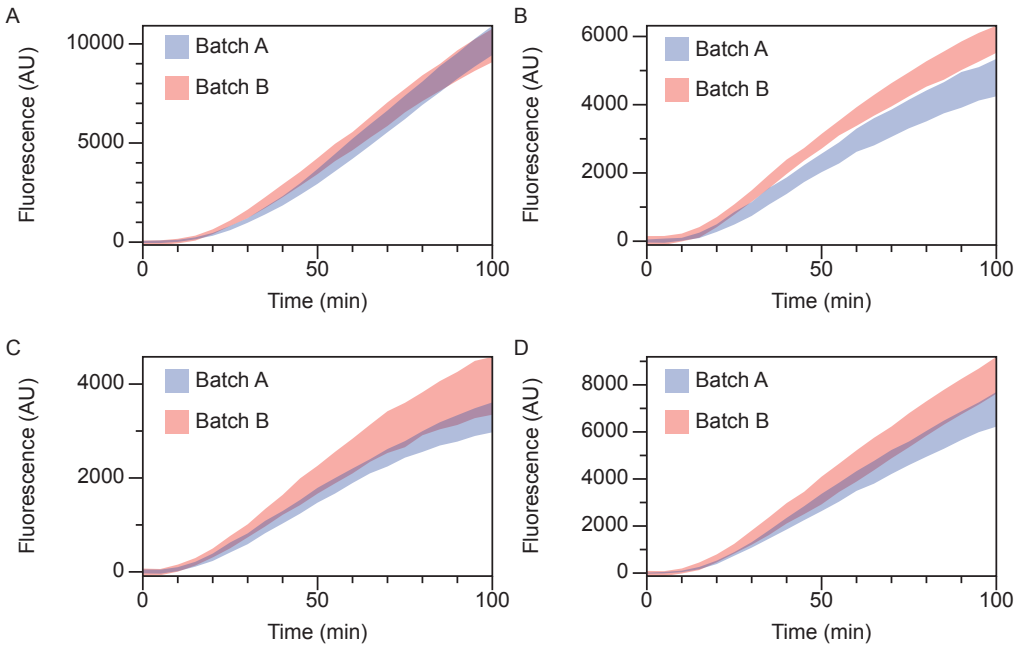


Figure A.7: Batch-to-batch difference Parameterization experimental trajectories comparison between batch A (blue) and batch B (red). Shaded regions represent the average plus/minus the standard deviation of experimental trajectories from Figure 2.4 (Batch A) and Figure A.5 (Batch B).

Table A.1: Important DNA sequences

Name	Sequence
pT181 attenuator (EcoRI-J23119-attenuator)	GAATTCTAAAGATCTTTGACAGCTAGCTCAGTCCTAGGTATAATAC TAGTAACAAAATAAAAAGGAGTCGCTCACGCCCTGACCAAAGTTT GTGAACGACATCATTCAAAGAAAAAAACACTGAGTTGTTTTTATA ATCTTGTATATTTAGATATTAAACGATATTTAAATATACATAAAGA TATATATTTGGGTGAGCGATTCCTTAAACGAAATTGAGATTAAGG AGTCGCTCTTTTTTATGTATAAAAACAATCATGCAAATCATTCAAA TCATTTGGAAAATCACGATTTAGACAATTTTTCTAAAACCGGCTAC TCTAATAGCCGGTTGTAAGGATCT

Table A.1 continued from previous page

pT181-mutant attenuator (EcoRI-J23119-attenuator)	GAATTCTAAAGATCTTTGACAGCTAGCTCAGTCCTAGGTATAATAC TAGTAACAAAATAAAAAGGAGTCGCTCTGTCCCTCGCCAAAGTTG CAGAACGACATCATTCAAAGAAAAAAACACTGAGTTGTTTTTATA ATCTTGTATATTTAGATATTAAACGATATTTAAATATACATAAAGA TATATATTTGGGTGAGCGATTCCTTAAACGAAATTGAGATTAAGG AGTCGCTCTTTTTTATGTATAAAAACAATCATGCAAATCATTCAAA TCATTTGGAAAATCACGATTTAGACAATTTTTCTAAAACCGGCTAC TCTAATAGCCGGTTGTAAGGATCT
Super folder green fluorescent protein (Ribosome binding site -SFGFP)	AGGAGGAAGGATCTATGAGCAAAGGAGAAGAACTTTTCACTGGA GTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACA AATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAA AACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCC GTGGCCAACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTT TCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTG CCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAG ATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTG ATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGA AGATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTC ACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCA AAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCA ACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCC TGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTT TCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAG TTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACA AATAA

Table A.1 continued from previous page

TrrnB	GAAGCTTGGGCCCGAACAAAAACTCATCTCAGAAGAGGATCTGAA TAGCGCCGTCGACCATCATCATCATCATCATTGAGTTTAAACGGTC TCCAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGAT ACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTT GCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAA CTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCC CCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAG GCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGG TGAACT
pT181 repressor	GAATTCTAAAGATCTTTGACAGCTAGCTCAGTCCTAGGTATAATAC TAGTATACAAGATTATAAAAACAACTCAGTGTTTTTTTCTTTGAAT GATGTCGTTCACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTA TTTGGATCT
pT181 mutant repressor	GAATTCTAAAGATCTTTGACAGCTAGCTCAGTCCTAGGTATAATAC TAGTATACAAGATTATAAAAACAACTCAGTGTTTTTTTCTTTGAAT GATGTCGTTCTGCAACTTTGGCGAGGGACAGAGCGACTCCTTTTTA TTTGGATCT
sTRSV Ribozyme	CTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGAGGACGAA ACAG

Table A.1 continued from previous page

Tandem pT181 attenuators	TTGACAGCTAGCTCAGTCCTAGGTATAATACTAGTAACAAAATAA AAAGGAGTCGCTCACGCCCTGACCAAAGTTTGTGAACGACATCAT TCAAAGAAAAAAACACTGAGTTGTTTTTATAATCTTGTATATTTAG ATATTAAACGATATTTAAATATACATAAAGATATATATTTGGGTGA GCGATTCCTTAAACGAAATTGAGATTAAGGAGTCGCTCTTTTTTAT GTATAAAAACAATCATGCAAATCATTCAAATCATTTGGAAAATCA CGATTTAGACAATTTTTCTAAAACCGGCTACTCTAATAGCCGGTTG TAACTCGAGAACAAAATAAAAAGGAGTCGCTCACGCCCTGACCAA AGTTTGTGAACGACATCATTCAAAGAAAAAAACACTGAGTTGTTT TTATAATCTTGTATATTTAGATATTAAACGATATTTAAATATACAT AAAGATATATATTTGGGTGAGCGATTCCTTAAACGAAATTGAGAT TAAGGAGTCGCTCTTTTTTATGTATAAAAACAATCATGCAAATCAT TCAAATCATTTGGAAAATCACGATTTAGACAATTTTTCTAAAACCG GCTACTCTAATAGCCGGTTGTAA
Double pT181 repressor	GGATCTCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGAG GACGAAACAGGGATCTATACAAGATTATAAAAACAACTCAGTGTT TTTTTCTTTGAATGATGTCGTTCACAAACTTTGGTCAGGGCGTGAG CGACTCCTTTTTATTTGGATCTCTGTCACCGGATGTGCTTTCCGGTC TGATGAGTCCGTGAGGACGAAACAGGGATCCTAACTCGAGATACA AGATTATAAAAACAACTCAGTGTTTTTTTCTTTGAATGATGTCGTT CACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTTGGATCT
Double pT181 mutant repressor	AGATGATCTCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT GAGGACGAAACAGGGATCTATACAAGATTATAAAAACAACTCAGT GTTTTTTTCTTTGAATGATGTCGTTCTGCAACTTTGGCGAGGGACA GAGCGACTCCTTTTTATTTCTCGAGCTGTCACCGGATGTGCTTTCC GGTCTGATGAGTCCGTGAGGACGAAACAGGGATCTATACAAGATT ATAAAAACAACTCAGTGTTTTTTTCTTTGAATGATGTCGTTCTGCA ACTTTGGCGAGGGACAGAGCGACTCCTTTTTATTTGGATCC

Table A.2: Plasmids used in this study. Sequences in the plasmid architecture can be found in Table A.1

Plasmid #	Plasmid architecture	Name	Figure
JBL006	J23119 pT181 attenuator SFGFP TrrnB CmR p15A origin	pT181 attenuator, Att-1 (Parameterization Cascade L1)	2.4,2.5,A.3,A.4, A.5,A.7
JBL002	J23119 TrrnB ColE1 origin AmpR	No repressor control	2.4,2.5,A.2,A.3, A.4.A.5,A.7
JBL007	J23119 pT181 mutant attenuator SFGFP TrrnB CmR p15A origin	pT181 mutant attenuator, Att-2 (Parameterization Cascade L2, Prediction Cascade PL1)	2.4,2.5,A.4,A.7
JBL008	J23119 pT181 mutant repressor TrrnB ColE1 origin AmpR	pT181 mutant repressor, AS-2 (Parameterization Cascade L3)	2.4,2.5,A.4,A.7
JBL069	J23119 pT181 mutant attenuator (sTRSV ribozyme pT181 repressor)x2 TrrnB ColE1 origin AmpR	Parameterization Cascade L2	2.4,2.5,A.2,A.3, A.4,A.7
JBL1845	J23119 (pT181 attenuator)x2 SFGFP TrrnB CmR p15A origin	Tandem pT181 attenuators, Att-1-Att-1 (Prediction Cascade L1T)	2.5,A.2
JBL3255	J23119 pT181 attenuator (sTRSV ribozyme pT181 mutant repressor)x2 TrrnB ColE1 origin AmpR	Prediction Cascade PL2	2.5
JBL004	J23119 pT181 repressor TrrnB ColE1 origin AmpR	pT181 repressor, AS-1 (Prediction Cascade PL3)	2.5,A.5,A.7

Table A.3: Estimation constraints

Parameter	Constrains
β_2	
β_1	
β_m	
K_2	$0.33\ K_1 \leqslant K_2 \leqslant 3.33K_1$
K_1	$0.33K_2 \leqslant K_1 \leqslant 3.33K_2$
k_E	$k_E \leqslant k_I$
d_2	$0.5d_1 \leqslant d_2 \leqslant d_1$
d_1	$2/3d_2 \leqslant d_1 \leqslant 2d_2$
d_m	$0.001 \leqslant d_m \leqslant d_2$
α	
k_I	$k_E \leqslant k_I \leqslant 10k_E$
r_{m2}	$0.33r_{m1} \leqslant r_{m2} \leqslant 3.33r_{m1}$
r_{m1}	$0.33r_{m2} \leqslant r_{m2} \leqslant 3.33r_{m2}$
p_c	$0.001 \leqslant p_c \leqslant 0.2$
P_t	$0.001 \leqslant P_t \leqslant 0.70$

Table A.4: Estimated parameter mean values

Parameter	Batch A	Batch B
β_m	2.0614 AU Conc. s^{-1}	3.1295 AU Conc. s^{-1}
K_2	132.02 AU Conc.	141.51 AU Conc.
K_1	240.89 AU Conc.	327.01 AU Conc.
k_E	0.0014572 s^{-1}	0.0013327 s^{-1}
d_2	0.0020798 s^{-1}	0.0039834 s^{-1}
d_1	0.00083283 s^{-1}	0.0015947 s^{-1}
d_m	0.0015946 s^{-1}	0.0032317 s^{-1}
α	0.01683 s^{-1}	0.049578 s^{-1}
k_I	0.002138 s^{-1}	0.002484 s^{-1}
r_{m2}	0.00015546 s^{-1}	0.00016851 s^{-1}
r_{m1}	3.2065e-05 s^{-1}	3.3942e-05 s^{-1}
p_c	0.020284	0.013106
P_t	0.0020295	0.001498

APPENDIX B

APPENDIX OF CHAPTER 3:ENGINEERING A FUNCTIONAL SMALL RNA NEGATIVE AUTOREGULATION NETWORK WITH MODEL-GUIDED DESIGN

B.1 Supplementary Note: ODEs that model individual parameterization experiments

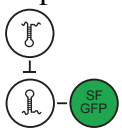
Experiment 1

$$\frac{dM}{dt} = \beta_m \cdot (1 - P_t) - d_M \cdot M \tag{1.1}$$

$$\frac{dG}{dt} = k_t \cdot M - \alpha \cdot G \tag{1.2}$$

$$\frac{dG_M}{dt} = \alpha \cdot G \tag{1.3}$$

Experiment 2



$$\frac{dR_1^*}{dt} = \beta_R - d_1 \cdot R_1^* - r_{m1} \cdot R_1^* \tag{2.1}$$

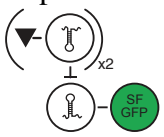
$$\frac{dR_1}{dt} = r_{m1} \cdot R_1^* - d_1 \cdot R_1 \tag{2.2}$$

$$\frac{dM}{dt} = \beta_m \cdot (1 - P_t) \cdot (1 - \frac{R_1}{K_1 + R_1}) - d_M \cdot M \tag{2.3}$$

$$\frac{dG}{dt} = k_t \cdot M - \alpha \cdot G \tag{2.4}$$

$$\frac{dG_M}{dt} = \alpha \cdot G \tag{2.5}$$

Experiment 3



$$\frac{dR_2^*}{dt} = \beta_R - d_2 \cdot R_2^* - r_{m2} \cdot R_2^* \tag{3.1}$$

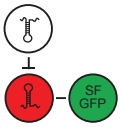
$$\frac{dR_2}{dt} = r_{m2} \cdot R_2^* - d_2 \cdot R_2 \tag{3.2}$$

$$\frac{dM}{dt} = \beta_m \cdot (1 - P_t) \cdot (1 - \frac{R_2}{K_2 + R_2}) - d_M \cdot M \tag{3.3}$$

$$\frac{dG}{dt} = k_t \cdot M - \alpha \cdot G \tag{3.4}$$

$$\frac{dG_M}{dt} = \alpha \cdot G \tag{3.5}$$

Experiment 4



$$\frac{dR_1^*}{dt} = \beta_R - d_1 \cdot R_1^* - r_{m1} \cdot R_1^* \tag{4.1}$$

$$\frac{dR_1}{dt} = r_{m1} \cdot R_1^* - d_1 \cdot R_1 \tag{4.2}$$

$$\frac{dM}{dt} = \beta_m \cdot (1 - P_t) \cdot (1 - \frac{R_1}{K_{C1} + R_1}) - d_M \cdot M \tag{4.3}$$

$$\frac{dG}{dt} = k_t \cdot M - \alpha \cdot G \tag{4.4}$$

$$\frac{dG_M}{dt} = \alpha \cdot G \tag{4.5}$$

Experiment 5

$$\frac{dMG^*}{dt} = \beta \cdot (1 - P_t) - d_{MG} \cdot MG^* - r_b \cdot MG^* \tag{5.1}$$

$$\frac{dMG}{dt} = r_b \cdot MG^* - d_{MG} \cdot MG \tag{5.2}$$

B.2 Methods for sensitivity analysis and parameter identifiability

Below we outline our methods for identifying parameters that can be fit from specifically designed experiments following2. Our model for RNA circuitry consists of a set of ordinary differential equations that describe the time-dependent rates of concentration changes of the molecular species that participate in the circuitry, $x_i(t)$. These equations are parameterized by a set of parameters, p_j, that we want to estimate by fitting model predictions to a small set of experiments. These experiments were designed through an iterative process of sensitivity analysis on the set of model equations (Figure B.4).

An individual experiment was considered to be a TX-TL reaction containing a subset of the DNA constructs encoding the full sRNA transcriptional NAR at defined concentrations. Each such experiment produces a measurable trajectory of SFGFP fluorescence as a function of time, and can be modeled by the subset of equations that describe the gene expression processes from the included DNA. After a specific experiment (a subset of DNA) was proposed, the next step was to assess which parameters were identifiable from this experiment, which is closely linked with parametric sensitivity analysis. Here we used the procedure proposed by McAuley and coworkers[127] to first calculate and analyze the sensitivity coefficient matrix for the proposed experiment as follows.

For each experiment, the sensitivity coefficient matrix $z_{ij}(t)$, is a time-varying matrix that encapsulates how sensitive the concentration of the molecular species x_i is to a change in the parameter p_j

$$Z_{ij}(t) = \left.\frac{\partial x_i}{\partial P_j}\right|_t \qquad i = 1, 2, ...N \qquad j = 1, 2, ..., P$$

Here P denotes the number of parameters and N denotes the number of molecular species. If we write the model equations generally as

$$\frac{dx_i}{dt} = f_i(\mathbf{x}, \mathbf{P}, t)$$

then it can be shown that $Z_{ij}(t)$ are the solutions to a set of differential equations given by

$$\frac{dz_{ij}}{dt} = \sum_{N}^{k=1} \frac{\partial f_i}{\partial x_k}(\mathbf{x}, \mathbf{P}, t) z_{k,j} + \frac{\partial f_i}{\partial p_j}(\mathbf{x}, \mathbf{P}, t)$$

which are subject to the initial condition $z_{ij} = 0$. Since our only observable in the TX-TL experiment is SFGFP, we focused specifically on $z_{SFGFP,j}(t)$ to determine which parameters were identifiable in the experiment.

Identifiability was then performed according to McAuley[127].This was done by finding the column of this matrix that had the biggest magnitude (indicating the most sensitive parameter), calculating a residual matrix which removed this column and controlled for correlations between parameters, and iterating this procedure on the resulting residual matrix until a threshold was reached on the largest remaining column magnitude. In this way a set of parameters was determined that maximally influenced the modeled trajectory of the proposed experiment (Figure 3.4).

After performing this procedure on the simplest experiment (experiment 1), we proposed a further experiment and performed the same analysis, except that parameters already identified by previous experiments were marked as determined by setting their columns in the sensitivity matrix to 0. Rounds of experimental design and sensitivity analysis were performed until all 16 parameters were able to be identified by five TX-TL experiments (Figure 3.4, Figure B.4).

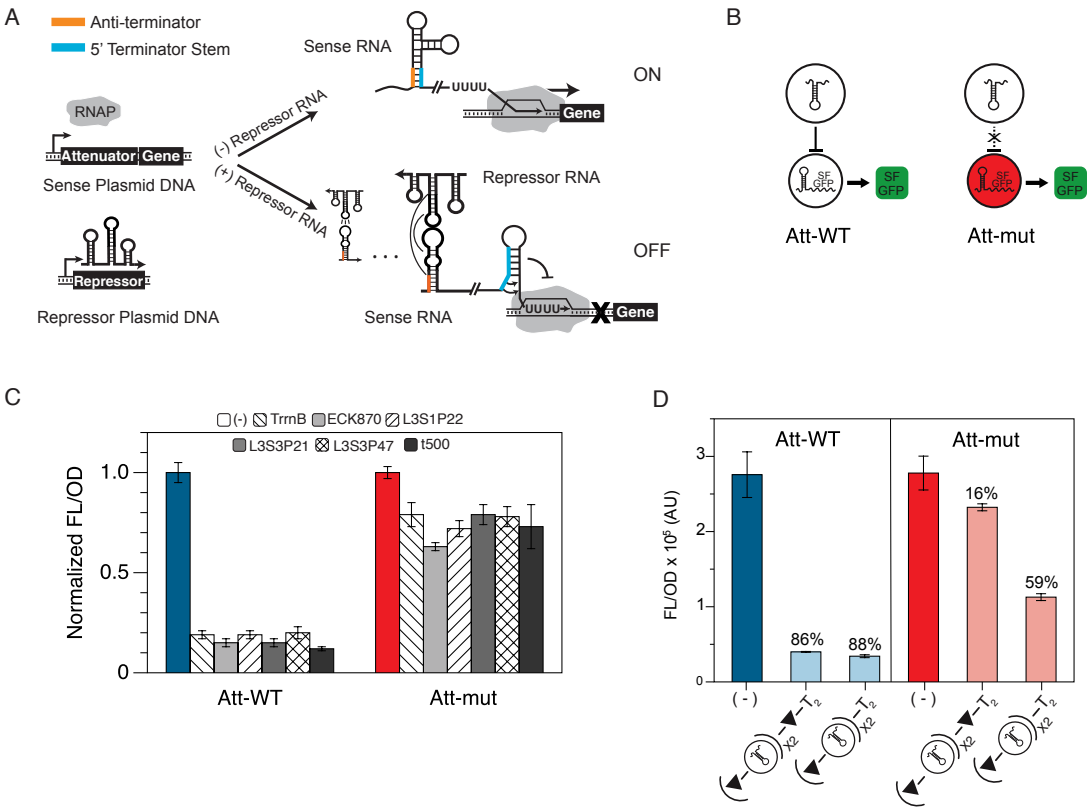


Figure B.1: (A) Natural repressor-RNA transcriptional repression in the pT181 attenuator. Transcription of a downstream gene is undisrupted when the repressor RNA is absent. When the repressor RNA is present, it interacts with the attenuator RNA to form a transcriptional terminator hairpin. Therefore, the transcription of the downstream gene is turned off. (B) Schematics of basic sRNA repression constructs testing both cognate and non-cognate (orthogonal) regulation. (C) In vivo test of 5 new terminators compared against TrrnB in the context of the sRNA repressor. Error bars represent standard deviations over nine biological replicates. L3S3P21 was chosen for the rest of this work. (D) *Invivo* orthogonality test of double (tandem) repressors. Adding a ribozyme (triangle) between the terminator and repressor construct restores orthogonality. Error bars represent standard deviations over three biological replicates.

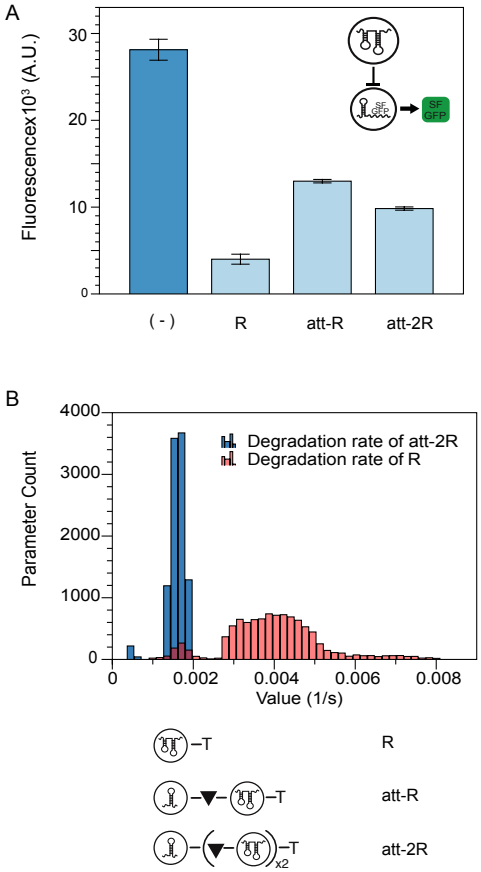


Figure B.2: Re-plots of previous findings on pT181 sRNA repression strength and degradation rates. (A) Re-plot of in vivo attenuator repression characterization from *Lucks et al*[64]. The repression efficiency of pT181 repressor antisense (R) is reduced when an attenuator sequence is placed upstream (att-R). Using two copies of repressors in tandem (att-2R) increases the repression strength slightly. (B) Re-plot of degradation parameter distributions for different repressor constructs determined from cell-free TX-TL experiments in *Hu et al*[48]. The degradation rate of the bare pT181 repressor sRNA (R, red) is faster than the degradation rate of the att-2R version (blue).

Table B.4: Important DNA sequences

Name	Sequence
ECK125109870 terminator	ccaattattgAACACCCTAACGGGTGTTTTTTTGTTTctggtctccc

Table B.4 continued from previous page

[-2em] L3S1P22 terminator	GACGAACAATAAGGCCGCAAATCGCGGCCTTTTTTATTGATAACAAAA
[-2em] L3S3P21 terminator	CCAATTATTGAAGGCCTCCCTAACGGGGGGCCTTTTTTTGTTTCTGGTCTCCC
[-2em] L3S3P47 terminator	TTTTCGAAAAAACACCCTAACGGGTGTTTTTTTATAGC TGGTCTCCC
Yem-GFP-LAA protein with degradation tag LAA	atgtctaaaggtgaagaattattcactggtgttgtcccaattttggttgaattagatggtgatgttaatggtcacaaattttctgtctccggtgaaggtgaaggtgatgctacttacggtaaattgaccttaaaatttatttgtactactggtaaattgccagttccatggccaaccttagtcactactttaacttatggtgttcaatgtttttctagatacccagatcatatgaaacaacatgacttttcaagtctgccatgccagaaggttatgttcaagaaagaactatttttttcaaagatgacggtaactacaagaccagagctgaagtcaagtttgaaggtgatacctagttaatagaatcgaattaaaaggtattgattttaaagaagatggtaacattttaggtcacaaattggaatacaactataactctcacaatgtttacatcatggctgacaaacaaaagaatggtatcaaagttaacttcaaaattagacacaacattgaagatggttctgttcaattagctgaccattatcaacaaaatactccaattggtgatggtccagtcttgttaccagacaaccattacttatccactcaatctaaattatccaaagatccaaacgaaaagagagaccacatggtcttgttagaatttgttactgctgctggtattacccatggtatggatgaattgtacaaaACTAGTGCAGCGAACGACGAAAATTACGCCCTTGCAGCG
T1 terminator	gcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgccctagac
p_{Lux} promoter	acctgtaggatcgtacaggtttacgcaagaaaatggtttgttatagtcgaataaa

Table B.4 continued from previous page

LuxR	atgaaaaacataaatgccgacgacacatacagaataattaataaaattaaagcttgtagaagcaataa tgatattaatcaatgcttatctgatatgactaaaatggtacattgtgaatattatttactcgcgatcatttatc ctcattctatggttaaatctgatatttcaatcctagataattaccctaaaaaatggaggcaatattatgatg acgctaatttaataaaatatgatcctatagtagattattctaactccaatcattcaccaattaattggaatat atttgaaaacaatgctgtaaataaaaaatctccaaatgtaattaaagaagcgaaaacatcaggtcttat cactgggtttagtttccctattcatacggctaacaatggcttcggaatgcttagttttgcacattcagaaa aagacaactatatagatagtttatttttacatgcgtgtatgaacataccattaattgttccttctctagttgat aattatcgaaaaataaatatagcaaataataaatcaaacaacgatttaaccaaaagagaaaaagaatg tttagcgtgggcatgcgaaggaaaaagctcttgggatatttcaaaaatattaggttgcagtgagcgta ctgtcactttccatttaaccaatgcgcaaatgaaactcaatacaacaaaccgctgccaaagtatttctaa agcaattttaacaggagcaattgattgcccatactttaaaaattaa
Malachite green aptamer (MG)	GGGATCCCGACTGGCGAGAGCCAGGTAACGAATGGATC
pT181 R(H2) single hairpin repressor	tctttgaatgatgtcgttcacaaactttggtcagggcgtgagcgactccttttattt
Mut pT181 R(H2), mutant single hairpin repressor	tctttgaatgatgtcgttcTGCaactttggCGagggACAgagcgactccttttattt
pT181-attenuator	AACAAAATAAAAAGGAGTCGCTCACGCCCTGACCAAAGTTT GTGAACGACATCATTCAAAGAAAAAAACACTGAGTTGTTTT TATAATCTTGTATATTTAGATATTAAACGATATTTAAATATA CATAAAGATATATATTTGGGTGAGCGATTCCTTAAACGAAA TTGAGATTAAGGAGTCGCTCTTTTTTATGTATAAAAACAATC ATGCAAATCATTCAAATCATTTGGAAAATCACGATTTAGAC AATTTTTCTAAAACCGGCTACTCTAATAGCCGGTTGTAA

Table B.4 continued from previous page

pT181-mutant attenuator	AACAAAATAAAAAGGAGTCGCTCTGTCCCTCGCCAAAGTTG CAGAACGACATCATTCAAAGAAAAAAACACTGAGTTGTTTT TATAATCTTGTATATTTAGATATTAAACGATATTTAAATATA CATAAAGATATATATTTGGGTGAGCGATTCCTTAAACGAAA TTGAGATTAAGGAGTCGCTCTTTTTTATGTATAAAAACAATC ATGCAAATCATTCAAATCATTTGGAAAATCACGATTTAGAC AATTTTTCTAAAACCGGCTACTCTAATAGCCGGTTGTAA
Super folder green fluorescent protein (Ribosome binding site - SFGFP)	AGGAGGAAGGATCTATGAGCAAAGGAGAAGAACTTTTCACT GGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAAT GGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGC TACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGG AAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGAC CTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAA ACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATG TACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTAC AAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGT TAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATG GAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCA CACAATGTATACATCACGGCAGACAAACAAAAGAATGGAAT CAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTT CCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTG GCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGA CACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGAC CACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACA CATGGCATGGATGAGCTCTACAAATAA

Table B.4 continued from previous page

TrrnB	GAAGCTTGGGCCCGAACAAAAACTCATCTCAGAAGAGGATC TGAATAGCGCCGTCGACCATCATCATCATCATCATTGAGTTT AAACGGTCTCCAGCTTGGCTGTTTTGGCGGATGAGAGAAGA TTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGT CTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGT CCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTA GCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGG AACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAG ACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACT
pT181 WT repressor	ATACAAGATTATAAAAACAACTCAGTGTTTTTTTCTTTGAAT GATGTCGTTCACAAACTTTGGTCAGGGCGTGAGCGACTCCTT TTTATTT
sTRSV Ribozyme	CTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGAGGA CGAAACAG
J23119 constitutive promoter	TTGACAGCTAGCTCAGTCCTAGGTATAATACTAGT

Table B.5: Plasmids used in this study. (Sequences in the plasmid architecture can be found in Table B.4)

Plasmid #	Plasmid architecture	Name	Figure
JBL3329	J23119-Att-yemGFP-LAA-T1 CmR/p15A	Yem-GFP	B.6A
JBL3349	pLuxR-pT181Att-yem-GFP-LAA-T1 CmR/p15A	pLux-Yem-GFP	B.6B
JBL3396	pLux-pT181Att-sTRSV-pT181R(H2)*-sTRSV-pT181 R(H2)-sTRSV-yemGFP-LAA	Double Repression NAR network in vivo	3.6, B.7

Table B.5 continued from previous page

JBL3398	pLuxR-pT181Attmut-sTRSV-pT181R(H2)-sTRSV-pT181R(H2)-sTRSV-yemGFP-LAA	Double Repression NAR CTRL network in vivo	3.6, B.7
JBL3368	pLux-pT181Att-sTRSV-pT181R(H2)-sTRSV-yemGFP-LAA	Single Repression NAR network in vivo	B.8
JBL3399	pLuxR-pT181Attmut-sTRSV-pT181R(H2)-sTRSV-yemGFP-LAA	Single Repression NAR CTRL network in vivo	B.8
JBL3343	J23119-sTRSVrbz-pT181R(H2)-sTRSVrbz-L3S3P21Term	Truncated pT181 repressor, parameterization experiment 2	3.4,3.2B,3.2A, B.6C
JBL5020	J23119-sTRSV-pT181R(H2)-sTRSV-pT181R(H2)-sTRSV-L3S3P21T	2Xtruncated pT181 repressor, parameterization experiment 3	3.4, B.1D
JBL3339	J23119-sTRSVrbz-pT181R(H2)-L3S3P21Term	Truncated pT181 repressor, no ribozyme before terminator	3.2B,B.2A
JB:5024	sTRSV-mutpT181R(H2)-sTRSV-P21T	Truncated pT181 mutated repressor	3.2B
JBL5025	sTRSV-mutpT181R(H2)-P21T	Truncated pT181 mutated repressor,no ribozyme before terminator	3.2B
JBL006	J23119-pT181att-SFGFP-TrrnB-CmRp15A origin	pT181 attenuator, Att-1 Parameterization experiment 1	3.4,B.1D, B.3B,B.6C
JBL002	J23119-TrrnB ColE1 origin AmpR	No repressor control	3.4, B.6C

Table B.5 continued from previous page

JBL007	J23119-pT181attmut-SFGFP-TrrnB-CmR p15A origin	pT181 mutant attenuator, parameterization experiment 4	3.4, B.3B, B.6C
JBL004	J23119-pT181R-TrrnB-ColE1 origin AmpR	pT181 repressor	3.2A
JBL1885	J23119-pT181R-ECK125109870term, ColE1 origin AmpR	Terminator variant	B.1C
JBL1886	J23119-pT181R-L3S1P22term ColE1 origin AmpR	Terminator variant	B.1C
JBL1887	J23119-pT181R-L3S3P21term ColE1 origin AmpR	Terminator variant	B.1C,3.2A
JBL1888	J23119-pT181R-L3S3P47term ColE1 origin AmpR	Terminator variant	B.1C
JBL3375	J23119-pT181Att-sTRSV-pT181R(H2)-sTRSV-MG-L3S3P21T	Single Repression NAR network TX-TL	3.5,B.7
JBL3376	J23119-pT181Att-sTRSV-pT181R(H2)-sTRSV-pT181R(H2)-sTRSV-MG-L3S3P21T	Double Repression NAR network TX-TL	3.5,B.7
JBL3377	J23119-pT181Attmut-sTRSV-pT181R(H2)-sTRSV-MG-L3S3P21T	Single Repression NAR CTRL network TX-TL	3.5,B.7
JBL3378	J23119-pT181Attmut-sTRSV-pT181R(H2)-sTRSV-pT181R(H2)-sTRSV-MG-L3S3P21T	Double Repression NAR CTRL network TX-TL	3.5,B.7

Table B.5 continued from previous page

JBL3358	J23119-pT181R(H2)-sTRSV-MG-L3S3P21T	R-MG	B.3A
JBL5032	J23119-MG-L3S3P21T	MG	B.3A
JBL3326	pT181-pT181R(H2)-L3S3P21 term	Truncated pT181 repressor	3.2A
JBL3355	J23119-sTRSV-pT181R(H2)-sTRSV-pT181R(H2)-L3S3P21T	2X truncated pT181 repressor, no ribozyme before terminator	3.2A, B.1D
JBL5054	J23119-pT181Attmut(H1)-sTRSV-pT181R(H2)-sTRSV-L3S3P21T	Mutated attenuator followed by 1 truncated pT181 repressor, with ribozyme before terminator	B.6C

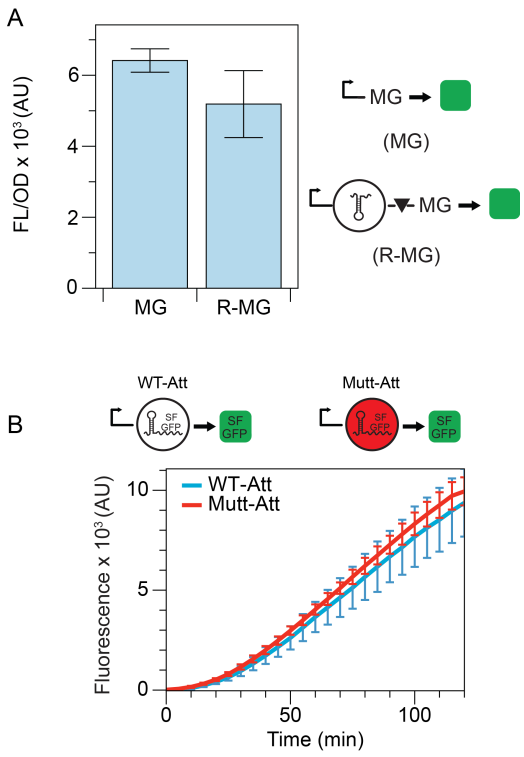


Figure B.3: Characterization of malachite green and SFGFP expression constructs. (A) End point fluorescent measurement at the end of a 2-hour TX-TL reaction with a constitutive promoter driving malachite-green aptamer expression (MG) compared with a version containing a repressor-ribozyme sequence in between the promoter and the MG aptamer (R-MG). This result indicates some reduction in MG transcription caused by the presence of the repressor-ribozyme sequence. The ribozyme is indicated by a triangle in the schematic. Error bars represent standard deviations of 3 technical replicates. (B) Two-hour TX-TL time course trajectories of constructs containing the wild type pT181 attenuator followed by an SFGFP coding sequence (WT-att), compared with a mutated version of the pT181 attenuator (Mutt-att) followed by an SFGFP coding sequence. The effective model simulates these two constructs with identical ODEs. Therefore, the two trajectories are expected to overlay each other if the DNA qualities of these two constructs are comparable. The result confirmed that this modeling approach is valid. Error bars represent standard deviations of 9 technical replicates.

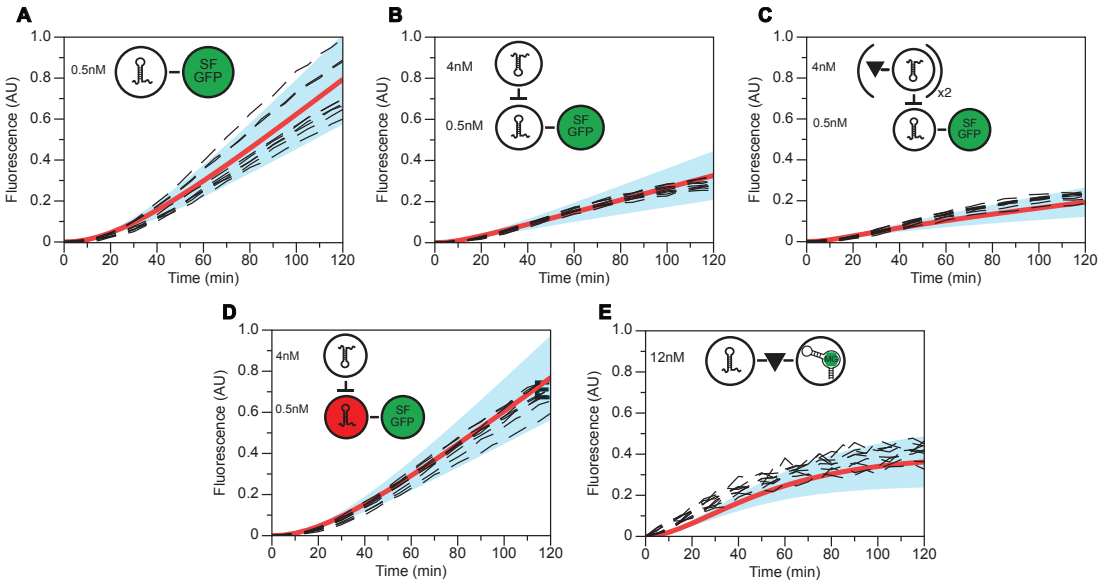


Figure B.4: Validation of model simulations of parameter estimation experiments. Comparison of experimental trajectories of SFGFP (or MG) fluorescence in TX-TL experiments (black dashed lines) with simulated model predictions. Model simulated trajectories were generated by performing 1000 simulations with parameters drawn from the set of 10,000 determined from the estimation procedure (see Methods). Experimental and model trajectories were normalized by the maximum observed experimental fluorescence of the first experiment in (A). The mean simulated trajectory (red line) is shown within 95% confidence intervals derived from the range of simulated trajectories (blue region). The schematic of each experiment is shown in the upper left corner of each plot corresponding to the experiments in Figure 3.4 and equations in Supplementary Note B.1.

Table B.1: Parameters guesses from previous work[48]

Parameter	Single-R	Tandem-R	Definition	Unit
1. β	1.5	1.5	Rate of transcription	AU Conc./sec
2. K	600	300	Repression co-efficient of RNA repressor	AU Conc.
3. d	0.008	0.004	Repressor degradation rate	1/sec
4. d_m	0.006	0.006	Reporter (G) degradation rate	1/sec

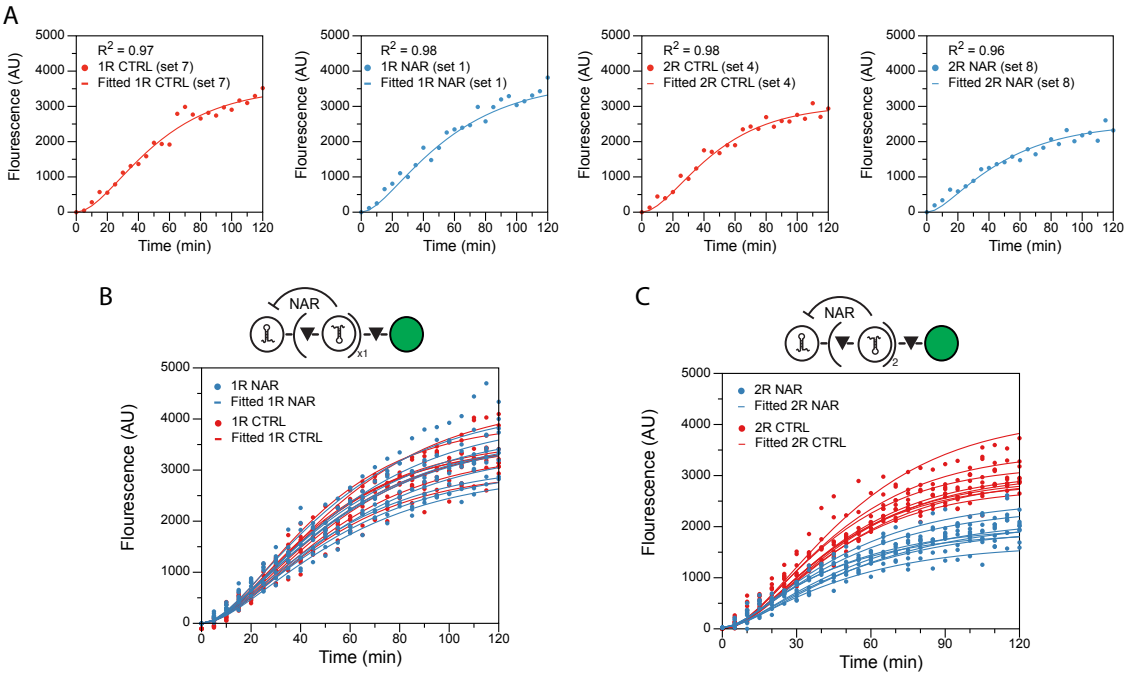


Figure B.5: To estimate the response time of dynamic trajectories generated by constructs in TX-TL reactions, we used a least squares method to approximate when each trajectory reaches steady state (see Methods). Since the raw data contains experimental noise within individual trajectories, steady state approximation is a difficult task. To overcome this, we fitted the data with our model and used the best fit trajectory to estimate the response time. (A) We searched for the best trajectory for each individual experimental replicate based on best R2 values. Sample fitting trajectories and experimental data from each experiment are shown. (B) All fitted trajectories and experimental data from the single repressor NAR vs. control constructs. (C) All fitted trajectories and experimental data from the double repressor NAR vs. control constructs. Response time of each experimental replicate (shown in Figure 3.5) was calculated independently following the least squared method (See Methods).

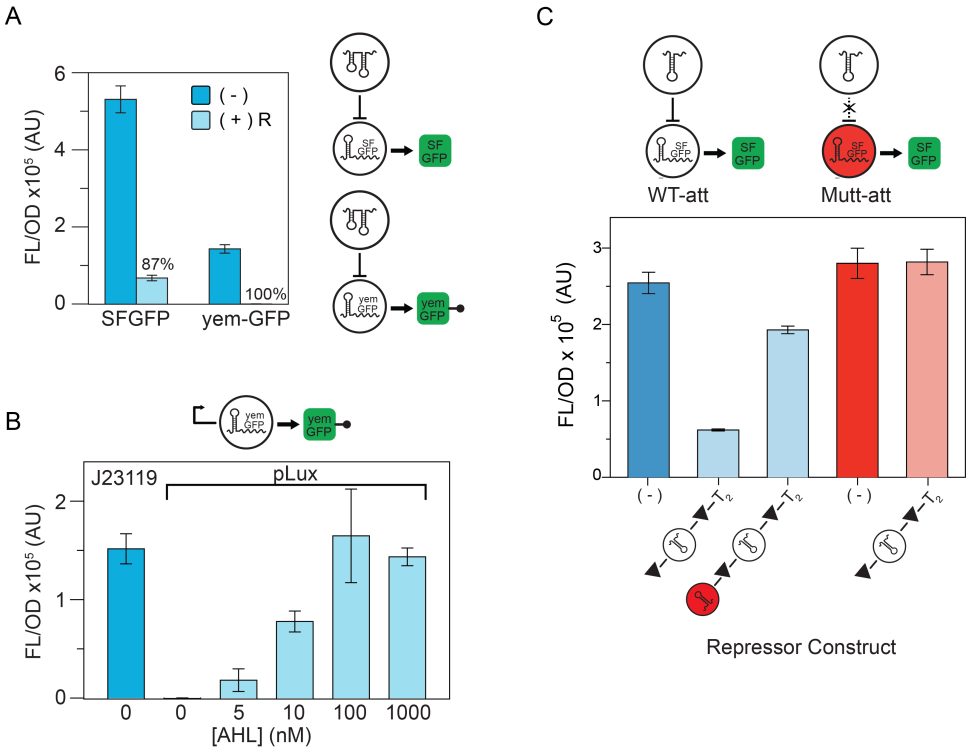


Figure B.6: Parts tested for the sRNA transcriptional NAR networks designed for in vivo experiments. (A) Testing yem-GFP in the context of the sRNA repressor. Expression was characterized from constructs containing a constitutive promoter followed by a pT181 attenuator and either GFP or yem-GFP with a control plasmid (-), or a plasmid encoding a cognate sRNA repressor (R,+). Results showed an improvement in repression strength when yem-GFP was used. Error bars represent standard deviations over nine biological replicates. (B) Using AHL to titrate expression of an attenuator construct. The attenuator-yem-GFP construct was placed behind the AHL inducible promoter pLux. Construct expression was measured after induction with a range of AHL concentrations after 5 hours, and compared to a constitutively expressed construct (J23119). The results confirmed a range of AHL induction levels of this repressor expression construct. Later in vivo experiments were performed with 100nM of AHL based on this result. Error bars represent standard deviations over nine biological replicates. (C) Repression efficiency of the single repressor construct used in the single repressor NAR networks. The repression strength of a single repressor flanked by ribozymes (triangle) versus the same construct with a mutant attenuator placed upstream. The present of the mutant attenuator significantly reduces the repression strength. Testing the repressor construct against a mutant attenuator target region (red bars) showed that the repression inefficiency was not due to crosstalk, but is rather due to the attenuator causing reduced sRNA transcription (see Figure B.2A). Error bars represent standard deviations over three technical replicates.

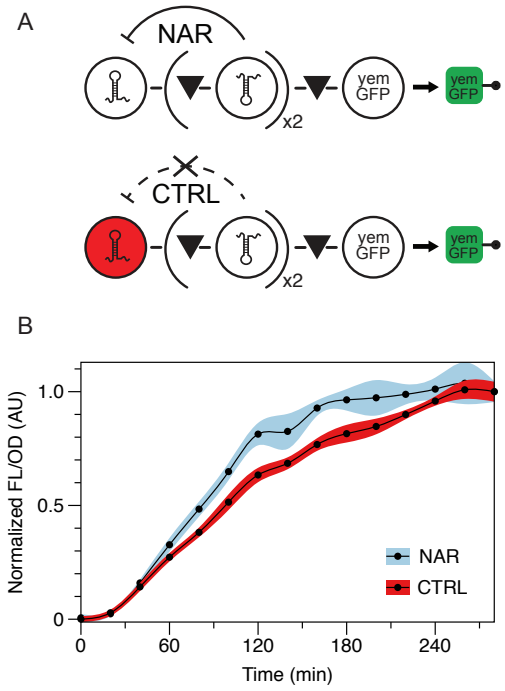


Figure B.7: Characterization of double repressor sRNA transcriptional NAR networks in vivo with normalized data. (A) Schematics of the double repressor NAR and control constructs designed for in vivo testing using yem-GFP as a network reporter. (B) Normalized trajectories collected from *E. coli* TG1 cells containing the double repressor NAR (blue) or control (red) construct over a five-hour period. Each experimental replicate from Figure 6B was normalized to its steady state value to show that NAR speeds up network response time.

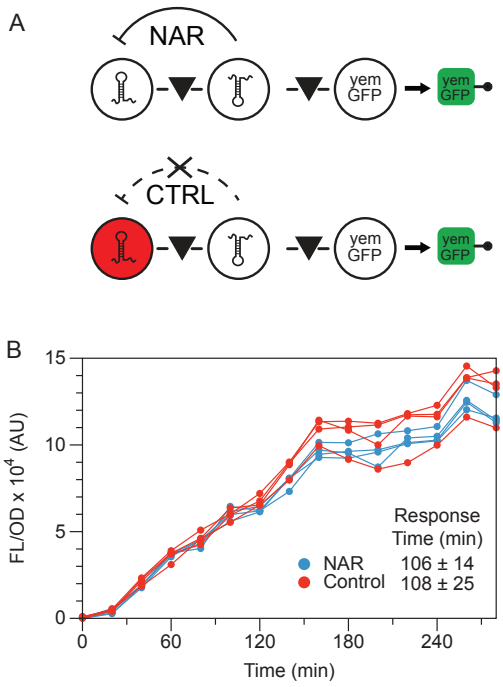


Figure B.8: Characterization of single repressor sRNA transcriptional NAR *in vivo* . (A) Schematics of the single repressor NAR and control constructs designed for *in vivo* testing using yem-GFP as a network reporter. (B) Four replicate fluorescence (FL/OD) trajectories collected from *E. coli* TG1 cells containing the single repressor NAR (blue) or control (red) construct over a five-hour period. In this case, no clear speed up in response was observed.

Table B.2: All parameters involved in parameterization procedure

#	Param.	Description	Estimated Value	Unit
P(1)	β_m	Transcription rate of mRNA	4.9 ± 0.41	Conc./sec.
P(2)	K_1	Repression coefficient of single repressor	289.5 ± 24.9	Conc.
P(3)	K_2	Repression coefficient of double (tandem) repressor	193.6 ± 16.7	Conc.
P(4)	K_{C1}	Repression coefficient of mismatched repressor due to crosstalk	10113.5 ± 886.0	Conc.
P(5)	d_1	Degradation rate of single repressor	2.1e-03 ± 2.8e-04	1/sec.
P(6)	r_{m1}	Maturation rate of single repressor	6.9e-05 ± 5.9e-06	1/sec.
P(7)	r_{m2}	Maturation rate of double repressor	7.5e-05 ± 6.4e-06	1/sec.
P(8)	r_b	Maturation and binding rate of MG	1.3e-04 ± 1.1-e05	1/sec.
P(9)	β_R	Transcription rate of repressor	39.4 ± 3.3	Conc./sec.
P(10)	k_t	Translational rate of SFGFP	2.6e-04 ± 2.3e-05	1/sec.
P(11)	α	Maturation rate of SFGFP	1.7e-02 ± 1.4e-03	1/sec.
P(12)	d_M	Degradation Rate of SFGFP mRNA	5.9e-04 ± 5.1e-05	1/sec.
P(13)	β	Transcription rate of pre-cleaved MG RNA	16.0 ± 1.4	Conc./sec.
P(14)	d_{MG}	Degradation rate of MG RNA	5.3e-04 ± 4.6e-05	1/sec.
P(15)	d_2	Degradation rate of double repressor	1.7e-03 ± 1.4e-04	1/sec.
P(16)	P_t	Auto-termination probability*	0.18 ± 0.016	N/A

Table B.3: Model Species

Model Species	Definition
R_1^*	Immature single sRNA repressor
R_1	Mature single sRNA repressor
R_2^*	Immature double (tandem) sRNA repressor
R_2	Mature double (tandem) repressor
M	mRNA of SFGFP
G	Immature SFGFP
G_M	Mature SFGFP
MG^*	Pre-cleaved R- MG
MG	Cleaved-off MG

APPENDIX C

APPENDIX OF CHAPTER 4: UNCOVER THE DESIGN PRINCIPLE OF GENERIC TRANSFER FUNCTION MANIPULATION THROUGH SMALL TRANSCRIPTION ACTIVATION RNAS (STARS)

C.1 Parametrization on Three Pairs of STARs

From previous work, we found that for the same circuit design, we would observe three distinct transfer functions and dynamic ranges for three different STAR/target pairs [23]. Intrigued by this observation, we aimed to gain better understanding of the mechanistic features governing STAR/target transfer functions through coarse grained ODE models. To estimate all unknown parameters, we proposed 6 TX-TL experiments to be performed sequentially [49]. Each experiment contained DNA that encoded parts of the attenuator circuitry that was described by a set of ODEs. The tables shown in below describe the species and parameter involved in STAR/target circuit.

C.1.1 Model Species and Parameters

Table C.1: Species involved in STAR/target Circuit

Species	Description
S^*	Immature STAR
S	Matured STAR
M	mRNA of GFP
G^*	Immature GFP
G	Matured GFP

Table C.2: Parameters involved in STAR/target Circuit

	#	Parameter	Description
FIXED	1	β_m	Transcriptional rate of mRNA
	2	d_m	Degradation rate of mRNA
	3	k_t	Translational rate of GFP
	4	α	Maturation rate of GFP
4970+4971	5	β_{s1}	Transcriptional rate of STAR
	6	d_{s1}	Degradation rate of STAR
	7	α_{s1}	Maturation rate of STAR
	8	P_1	Autotermination 0 – 1.0
	9	K_{s1}	Activation coefficient of STAR
	10	l_{s1}	Leak coefficient of STAR 0 – 1.0
5808+5809	11	β_{s2}	Transcriptional rate of STAR
	12	d_{s2}	Degradation rate of STAR
	13	α_{s2}	Maturation rate of STAR
	14	P_2	Autotermination 0 – 1.0
	15	K_{s2}	Activation coefficient of STAR
	16	l_{s2}	Leak coefficient of STAR 0 – 1.0
5816+5817	17	β_{s3}	Transcriptional rate of STAR
	18	d_{s3}	Degradation rate of STAR
	19	α_{s3}	Maturation rate of STAR
	20	P_3	Autotermination 0 – 1.0
	21	K_{s3}	Activation coefficient of STAR
	22	l_{s2}	Leak coefficient of STAR 0 – 1.0

C.1.2 Equations

$$\frac{dS^*}{dt} = \beta_{s,i} - d_{s,i}S^* - \alpha_{s,i}S^* \quad \text{(C.1)}$$

$$\frac{dS}{dt} = \alpha_{s,i}S^* - d_{s,i}S \quad \text{(C.2)}$$

$$\frac{dM}{dt} = \beta_m(1 - P_i)\left(\frac{S}{K_{s,i} + S} + l_{s,i}\right) - d_m M \quad \text{(C.3)}$$

$$\frac{dG^*}{dt} = k_t M - \alpha G^* \quad \text{(C.4)}$$

$$\frac{dG}{dt} = \alpha G^* \quad \text{(C.5)}$$

where $i = 1, 2, 3$ denotes 4970/4971 pair, 5808/5809 pair, and 5816/5817 pair, respectively. In the parametrization process, construct levels 1 and 2 correspond to experiments performed with 4970/4971 STAR/target pair. The first construct level contains only a set concentration of the 4970 target DNA, representing the off level of the 4970/4971 STAR/target pair. The second construct level contains both the 4971 STAR DNA and the 4970 target DNA at set concentrations, representing the on level of the 4970/4971 STAR/target pair. Likewise, construct levels 3 and 4 contain the 5808 target DNA only and the 5808/5809 STAR/target DNAs, respectively. Construct levels 5 and 6 contain the 5816 target DNA only and the 5816/5817 STAR/target DNAs, respectively. For each off level (construct levels 1,3, or 5), parameter $\beta_{s,i}$ corresponding to each construct level is set to zero to represent the absence of STAR in each system.

C.1.3 Parametrization Methods

Our model describing the STAR RNA circuitry consists of a set of ordinary differential equations that describes the time varying rate of change in the concentrations of each molecular species, $x_i(t)$. These equations are parametrized by a set of parameters, p_j, that are estimated by fitting model predictions to a small set of experiments [49]. The

experiments are designed through an iterative processes of sensitivity analysis on the set of model equations until all parameters are found.

A TX-TL reaction containing a subset of the DNA constructs encoding the full STAR and Target DNA construct at defined concentrations is considered to be an individual experiment. Each such experiment produces a measurable trajectory of SFGFP fluorescence as a function of time. The trajectory obtained can in turn be modeled by a subset of equations that describes the gene expression processes from the included DNA. After a specific experiment was proposed, we assessed which parameters can be "identified" from such experiment. This identifiability assessment is closely linked with parametric sensitivity analysis. The sensitivity analysis used here are proposed by McAuley and coworkers [128] to calculate and analysis the sensitivity coefficient matrix for each experiment.

Each proposed experiment has a time varying sensitivity coefficient matrix $z_{ij}(t)$ that captures how sensitive the concentration of each molecular species x_i is to a change in the parameter p_j:

$$z_{ij} = \left.\frac{\partial x_i}{\partial p_j}\right|_t \tag{C.6}$$

with $i = 1, 2, ..., N$ and $j = 1, 2, ...P$, where N denotes the number of parameters and N denotes the number of molecular species.

The model equations can be generically expressed as:

$$\frac{dx_i}{dp_j} = f_i(\vec{x}, \vec{p}, t) \tag{C.7}$$

and $z_{ij}(t)$ are the solutions to a set of differential equations given by

$$\frac{dz_{i,j}}{dt} = \sum_{k=1}^{N} \frac{\partial f_i}{\partial x_k}(\vec{x}, \vec{p}, t) z_{k,j} + \frac{\partial f_i}{\partial p_j}(\vec{x}, \vec{p}, t) \tag{C.8}$$

subjected to the initial condition $z_{ij} = 0$. Since the only measurable specie in our proposed experiment is SFGFP, we focused on $z_{SFGFP,j}(t)$ to determine which parameter is identifiable in the experiment.

Identifiability assessment was then performed according to procedures proposed by McAuley [128] - first find the column of the sensitivity matrix that has the largest magnitude (most sensitive parameter), and then compute a residual matrix which removed this column and controlled for correlations between parameters. This procedure is repeated on resulting residual matrix until a threshold was reached on the largest remaining column magnitude. The result is a determined set of parameters that will maximally influence the modeled trajectory of the proposed experiment.

After performing this method on the first experiment, we proposed subsequent experiments and conducted the same analysis. The parameters already identified by previous experiments were marked as ”determined” by setting their columns in the sensitivity matrix to zero. Experimental design and sensitivity analysis were repeated until all parameters could be identified by TX-TL experiments.

C.1.4 Sensitivity Analysis

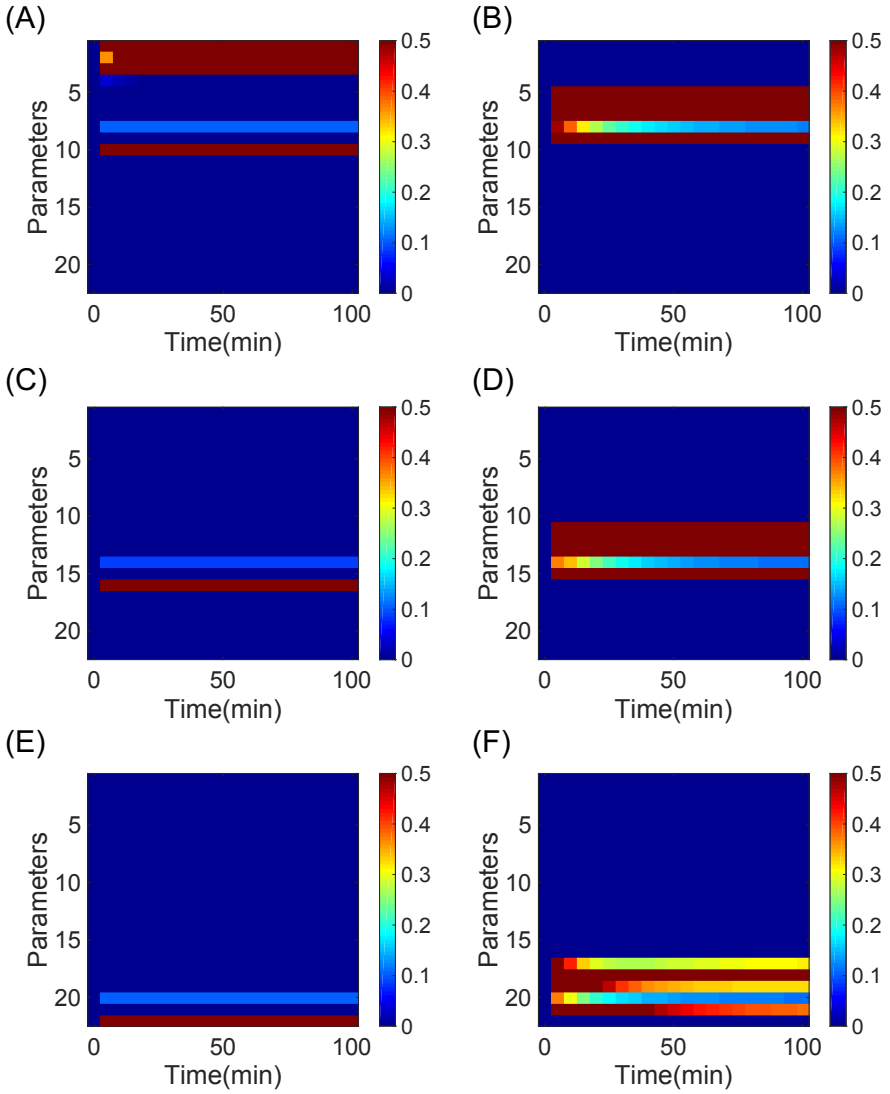


Figure C.1: Six parametrization experiments were designed on the basis of sensitivity analysis to estimate all 22 unknown parameter values in our proposed model. Red and blue indicated high and low sensitivity, respectively. The first construct level (A) contains only a set concentration of the 4970 target DNA, representing the off level of the 4970/4971 STAR/target pair. The second construct level (B) contains both the 4971 STAR DNA and the 4970 target DNA at set concentrations, representing the on level. Likewise, construct levels 3 (C) and 4 (D) contain the 5808 target DNA and the 5808/5809 STAR/target DNAs, respectively. Construct levels 5 (E) and 6 (F) contain the 5816 target DNA only and the 5816/5817 STAR/target DNAs, respectively.

C.1.5 Simulation Results

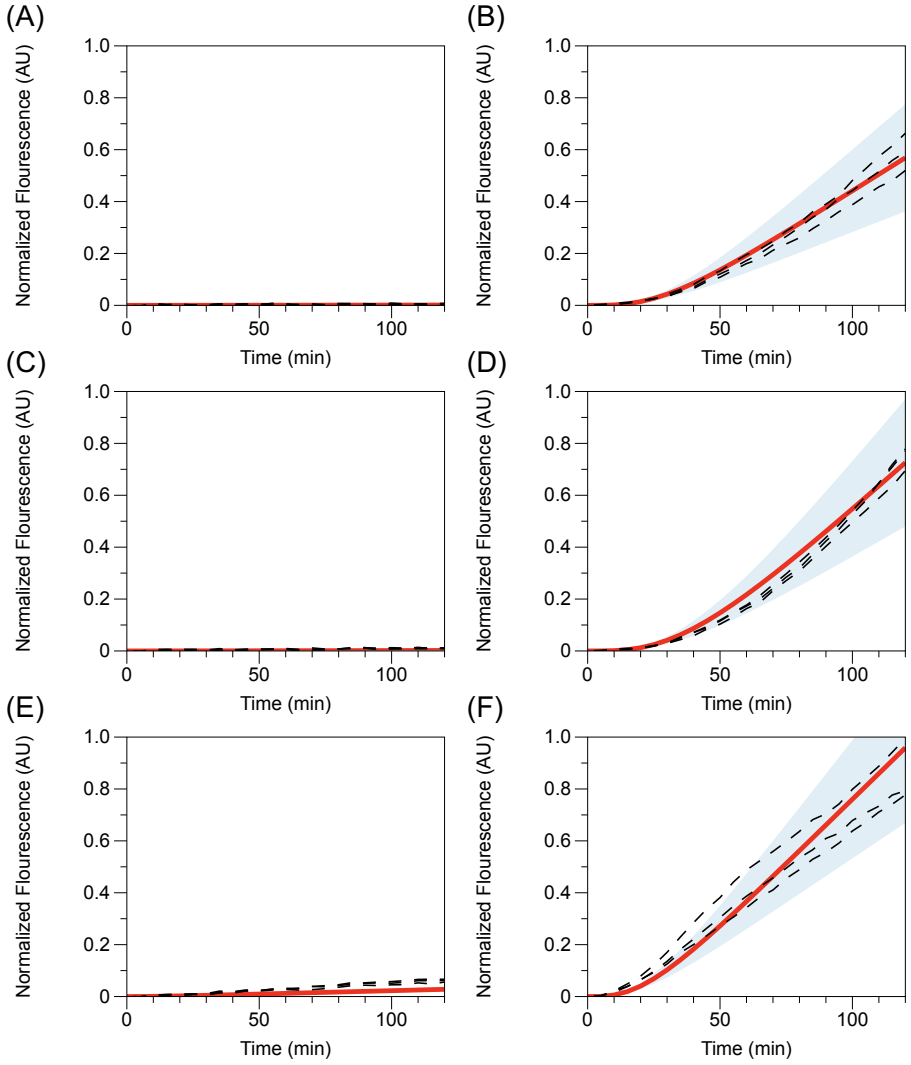


Figure C.2: Comparison of experimental trajectories of SFGFP fluorescence in TX-TL experiments (black dashed lines) with simulated model predictions. Model simulated trajectories were generated by performing 100 simulations with parameters drawn from the set of 10,000 determined from the parameter estimation procedure. Experimental and model trajectories were normalized by the maximum experimental fluorescence of the first experiment (A). The mean simulated trajectory (red line) is shown within 95% confidence intervals (shaded blue region). Plots (A) and (B) 2 correspond to 4970 and 4971 pair. Plots (C) and (D) correspond to 5808 and 5809 pair. Plots (E) and (F) correspond to 5816 and 5817 pair.

C.1.6 STAR Pairs and Parametrized Values

Table C.3: Estimated Parameter Values

	#	Parameter	Mean Values	Std.dev
FIXED	1	β_m	1.4587	0.12801
	2	d_m	0.001284	0.00010484
	3	$k_t l$	0.0048837	0.00041862
	4	α	0.15091	0.013875
4970+4971	5	β_s	4.3079	0.38497
Best OFF	6	d_s	0.0015914	0.00014331
	7	α_s	5.98E-05	5.45E-06
	8	P	0.068187	0.0062317
	9	K_s	199.29	17.09
	10	l_s	0.00058557	5.52E-05
5808+5809	11	β_s	3.5585	0.29702
Best Dynamic	12	d_s	0.00093809	7.64E-05
Range	13	α_s	2.14E-05	1.84E-06
	14	P	0.12277	0.010304
	15	K_s	89.335	7.7044
	16	l_s	0.00094246	8.68E-05
5816+5817	17	β_s	4.0062	0.3572
Best ON	18	d_s	0.0022392	0.00018976
	19	α_s	0.00014178	1.24E-05
	20	P	0.14601	0.012404
	21	K_s	86.875	7.8063
	22	l_s	0.014516	0.0012806

Table C.3 shows the mean values of estimated parameters as well as standard deviations.

These estimated parameter values agreed with previous experiments [23]. For instance, we previously observed the 4970/4791 STAR/target pair have the lowest leak in off state and the lowest signal at on state among the three pairs of STAR/target tested *in vivo*. The estimated values of parameters corresponding to these features, l_s and K_s are the lowest and the highest among the three pairs of STAR/target. Likewise, 5816/5817 STAR/target pair has the highest leak in off state but also the highest signal at on state. The estimated values of parameters corresponding to these features, l_s and K_s are the highest and the lowest among the three pairs of STAR/target.

C.2 *In vivo* Simulation with Three Pairs of STARs

To test the robustness of the parameter values estimated from parameterization experiments, we aimed to simulate the *in vivo* transfer functions of hybrid inducible systems with three pairs of STARs. In previous work, we have shown that parameter values extracted from TX-TL experiments can serve as prototyping tools for characterizing circuit behaviors *in vivo* [47]. The tables below describe the species and parameters used in *in vivo* model.

C.2.1 Model Species and Parameters

Table C.4: Species involved in hybrid inducible system

Species	Description
S	STAR
M	mRNA of GFP
G	GFP

C.2.2 *In vivo* Equations

The ODE describing the maturation step of a newly transcribed STAR (S^*) into a functional STAR (S) is:

$$\frac{dS}{dt} = \alpha_{s,i} S^* - d_{s,i} S \tag{C.9}$$

For *in vivo* experiment, we only care about the concentration profiles of three species: STAR (S), mRNA of GFP (M), and GFP (P). To recapitulate the maturation step of S, we can set the above expression to steady state and obtain the scaling relation between an immature STAR and a matured STAR:

$$S_{matured} = \frac{\alpha_{s,i}}{d_{s,i}} S_{immature} \tag{C.10}$$

Table C.5: Parameters involved in hybrid inducible system

Parameter	Description
X	Inducer Concentration
K_x	Activation coefficient of Inducer
l_x	Leak coefficient of Inducer
β_x	Transcriptional rate of STAR driven by inducible promoter
d_s	Degradation rate of STAR
α_s	Maturation rate of STAR
β_m	Transcriptional rate of mRNA driven by constitutive promoter
K_s	Activation coefficient of STAR
d_m	Degradation rate of mRNA
k_t	Translational rate of GFP
d_p	Degradation rate of GFP
l_s	Leak coefficient of STAR (0 – 1.0)
P	Autotermination

This scaling relation (α_s/d_s) is used in the hill function for mRNA to account for the amount of matured STAR capable of activating GFP expression:

$$\frac{dS}{dt} = \beta_s\left(\frac{X}{K_x + X} + l_x\right) - d_s S \tag{C.11}$$

$$\frac{dM}{dt} = \beta_m(1 - P)\left(\frac{\alpha_s S/d_s}{K_s + \alpha_s S/d_s} + l_s\right) - d_m M \tag{C.12}$$

$$\frac{dG}{dt} = k_t M - d_p G \tag{C.13}$$

Table C.6: Parameters values used in hybrid inducible system

Parameter	Values	Units
X	$0.001 - 1 \times 10^6$	molecules
K_x	20	molecules
l_x	0.05	molecules
β_x	2	1/sec
d_s	drawn from parameter pool	1/sec
α_s	drawn from parameter pool	1/sec
β_m	4	1/sec
K_s	drawn from parameter pool	molecules
d_m	drawn from parameter pool	1/sec
k_t	0.034 [41]	1/sec
d_p	0.0001 [2]	1/sec
l_s	Drawn from parameter pool	N/A
P	Drawn from parameter pool	N/A

C.2.3 Parameter Values used for *in vivo* Simulation

In the *in vivo* model simulation, we have drawn several STAR parameters values from the estimation pool (d_s, α_s, K_s, d_m) in combination with literature values of protein translation k_t and degradation d_p speeds. For parameter such as $X, K_x, l_x, \beta_x, \beta_m$ where literature values are not available, we provided reasonable estimates based on previous works.

C.3 Understanding Leak coefficient l_x of transcription factor mediated activator

The following ordinary differential equations describe the gene expression event of a protein P controlled by an activating transcriptional regulator that responses to inducer x.

$$\frac{dM}{dt} = \beta \cdot (\frac{x}{K_x + x} + l_x) - d_m \cdot M \quad \text{(C.14)}$$

$$\frac{dP}{dt} = k_t \cdot M - d_p \cdot P \quad \text{(C.15)}$$

Equations C.14 and C.15 represent the transcriptional and translational events in gene expression. Here M denotes the concentration of mRNA and P denotes for the concentration of protein. β and k_t denote the rate of transcription and translation, respectively. d_m and d_p denote the degradation rate of mRNA and protein, respectively. K_x is the activation strength and l_x represents the leak coefficient, which ranges from 0 to 1. It defines the amount of mRNA being expressed when the system is not being induced. The causes of leak differ from systems to systems. Here we explore the cause of leak on several common inducible systems.

C.3.1 Positive Activation

First of all, a transcriptional factor (TF) mediated activator could be categorized into two groups based on its activating mechanism – positive activation and repressor sequestration. A positive activation mediated by TF can be broken down int o two steps: step one, inducer x binding with transcriptional factor TF to form activating complex C:

$$\mathrm{x} + \mathrm{TF} \underset{k_{1r}}{\overset{k_{1f}}{\rightleftharpoons}} \mathrm{C}$$

$$\frac{dC}{dt} = k_{1f} \cdot x \cdot TF - C \cdot k_{1r}$$

At steady state and equilibrium,

$$C = \frac{k_{1f}}{k_{1r}} \cdot x \cdot TF$$

If $x_{tot} = x + C$, and $k_{1D} = \frac{k_{1r}}{k_{1f}}$ Then

$$C = x_{tot} \cdot \frac{TF}{k_{1D} + TF} \tag{C.16}$$

Step two, activating complex C binds to promoter region p and turn it into an active transcript T_a,

$$\mathrm{C} + \mathrm{p} \underset{k_{2r}}{\overset{k_{2f}}{\rightleftharpoons}} \mathrm{T_a}$$

$$\frac{dT_a}{dt} = k_{2f} \cdot C \cdot p - T_a \cdot k_{2r}$$

At steady state and equilibrium,

$$T_a = \frac{k_{2f}}{k_{2r}} \cdot C \cdot p$$

If $p_{tot} = T_a + p$, and $k_{2D} = \frac{k_{2r}}{k_{2f}}$ Then

$$T_a = p_{tot} \cdot \frac{C}{k_{2D} + C} \tag{C.17}$$

Sometimes off-target interactions would cause leak: For instance, transcriptional factor TF activates promoter P into an active transcript T_a:

$$\mathrm{TF} + \mathrm{P} \underset{k_{1lr}}{\overset{k_{1lf}}{\rightleftharpoons}} \mathrm{T_a}$$

If $k_{1lD} = \frac{k_{1lr}}{k_{1lf}}$

$$T_a = p_{tot} \cdot \frac{TF}{k_{1lD} + TF} \tag{C.18}$$

When the promoter binds to RNA polymerase and turns into an active transcript:

$$\mathrm{P} \underset{k_{2lr}}{\overset{k_{2lf}}{\rightleftharpoons}} \mathrm{T_a}$$

If $k_{2lD} = \frac{k_{2lr}}{k_{2lf}}$

$$T_a = \frac{p_{tot}}{k_{2lD}} \tag{C.19}$$

If we combine equations C.16 to C.19 and substitute them into equation C.14

$$\frac{dM}{dt} = f_{TF} \cdot p_{tot}(\frac{x_{tot}}{k_{2D} + x_{tot}} + l_x) - d_m M \tag{C.20}$$

where

$$f_{TF} = \frac{TF}{k_{1D} + TF}$$

$$l_x = \frac{TF}{k_{1lD} + TF} + \frac{1}{k_{2lD}}$$

Therefore, in order to minimize the leak l_x, the ideal solution would be to engineer the transcriptional factor TF to avoid non-specific binding with the promoter, therefore increase k_{1lD}; also, to engineer the promoter sequence to avoid self-promoting, therefore increase k_{2lD}. However, these are non-trivial engineering tasks. Different system requires specific mechanistic tuning on the molecular level to solve this problem.

C.3.2 Repressor sequestrating activation

In repressor sequestration activation, inducer x binds to the transcription factor TF and forms complex C_1. This allows the transcription factor to be released from the promoter and de-represses transcription:

$$\mathrm{x} + \mathrm{TF} \underset{k_{1r}}{\overset{k_{1f}}{\rightleftharpoons}} \mathrm{C_1}$$

$$\frac{dC_1}{dt} = k_{1f} \cdot x \cdot TF - C_1 \cdot k_{1r}$$

At steady state and equilibrium,

$$C_1 = \frac{k_{1f}}{k_{1r}} \cdot x \cdot TF$$

If $TF_{tot} = TF + C_1$, and $k_{1D} = \frac{k_{1r}}{k_{1f}}$, we have

$$TF = TF_{tot} \cdot \frac{k_{1D}}{x + k_{1D}} \tag{C.21}$$

The active transcript is repressed by the transcription factor TF by the following mechanism.

$$\mathrm{T_a + TF} \underset{k_{2r}}{\overset{k_{2f}}{\rightleftharpoons}} \mathrm{C_2}$$

$$\frac{dC_2}{dt} = k_{2f} \cdot T_a \cdot TF - C_2 \cdot k_{2r}$$

At steady state and equilibrium:

$$C_2 = \frac{k_{2f}}{k_{2r}} \cdot T_a \cdot TF$$

If $T_{a,tot} = T_a + C_2$, and $k_{2D} = \frac{k_{2r}}{k_{2f}}$, we have

$$T_a = T_{a,tot} \cdot \frac{k_{2D}}{TF + k_{2D}} \tag{C.22}$$

where total number of active transcript $T_{a,tot}$ is given by the total number of promoter P_{tot} in the system. We then plug in the expression of TF in eqn.C.21 into eqn.C.22 and obtain:

$$T_a = P_{tot} \frac{x + k_{1D}}{TF_{tot} \dfrac{k_{1D}}{k_{2D}} + x + k_{1D}} \tag{C.23}$$

With no inducer in the system, the leak l_x is given by the amount of active transcript T_a with $x = 0$:

$$l_x = P_{tot} \frac{1}{1 + \dfrac{TF_{tot}}{k_{2D}}} \tag{C.24}$$

In order to minimize leak l_x, we can increase the amount of repressors (TF_{tot}) in the system. Another approach is to engineer the repressor to have tight binding with the promoter, thereby decreasing k_{2D}. Tuning the right amount of repressors in the system is non-trival: having too much repressors or a low k_{2D} where $TF_{tot}\frac{k_{1D}}{k_{2D}} \gg x + k_{1D}$ in eqn.C.24 makes the system non-inducible.

C.4 Degradation Model for sRNA

Experiments showed that hybrid-inducible system with STAR intermediate improves dynamic range compared to inducible system alone. However, our original model showed that, while hybrid system can reduce off level leak, it should not improve dynamic range. To account for the discrepancy in experimental observations and model, we incorporated degradation steps into the original model. The new degradation model accounts for the dynamic recruitment of RNase at on and off levels in the hybrid system. We hypothesize that the observable degradation rate is faster at off level, as there are more RNase than sRNAs for efficient degradation. At on level, more sRNAs are transcribed and saturated RNase activity, the observable degradation rate would be slower. Tables below detail the model species and parameters used:

C.4.1 Model Species and Parameters

Table C.7: Species involved in STAR-hybrid system

Species	Description
S	STAR
M	mRNA of GFP
Cd_1	sRNA-RNase Complex
Cd_2	mRNA-RNase Complex
A	Total RNase
G	GFP

Table C.8: Parameters involved in STAR-hybrid system

Parameter	number	Description	Value
X	1	Inducer Concentration	0.001 - 1e6
K_x	2	Activation coefficient of Inducer	20
l_x	3	Leak coefficient of Inducer	0.25
β_m	4	Transcriptional rate driven by constitutive promoter	4
d_s	5	Degradation rate of STAR	0.002
β_x	6	Max transcriptional rate driven by inducible promoter	2
K_s	7	Activation coefficient of STAR	2000
d_m	8	Degradation rate of mRNA	0.01
k_t	9	Translational rate of GFP	0.035
d_p	10	Degradation rate of GFP	0.0001
l_s	11	Leak coefficient of STAR ($0-1.0$)	0.001
k_{f1}	12	Forward binding constant of sRNA to RNase	1
k_{r1}	13	Reverse binding constant of sRNA to RNase	1000
k_{f2}	14	Forward binding constant of mRNA to RNase	1
k_{r2}	15	Reverse binding constant of mRNA to RNase	1000

C.4.2 Equation

Here we constructed three models: inducible system alone, original hybrid system model with constant degradation, and new hybrid system model accounting RNase saturation. The same parameter values are used in simulating the dynamic range for all three models.

Without STAR intermediate

$$\frac{dM}{dt} = \beta_m\left(\frac{X}{K_x + X} + l_X\right) - d_m M \tag{C.25}$$

$$\frac{dG}{dt} = k_t M - d_p G \tag{C.26}$$

Constant Degradation

$$\frac{dS}{dt} = \beta_x\left(\frac{X}{K_x + X} + l_x\right) - d_s S \tag{C.27}$$

$$\frac{dM}{dt} = \beta_m\left(\frac{S}{K_s + S} + l_S\right) - d_m M \tag{C.28}$$

$$\frac{dG}{dt} = k_t M - d_p G \tag{C.29}$$

Changing Degradation

In the degradation model, the dynamics of RNase activities are accounted for by the proposed mechanisms:

$$S + A \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} Cd_1 \xrightarrow{d_s} A + \emptyset$$

$$M + A \underset{k_{r2}}{\overset{k_{f2}}{\rightleftharpoons}} Cd_2 \xrightarrow{d_m} A + \emptyset$$

$$\frac{dS}{dt} = \beta_x\left(\frac{X}{K_x + X} + l_x\right) - k_{f1}SA + k_{r1}Cd_1 \tag{C.30}$$

$$\frac{dM}{dt} = \beta_m\left(\frac{S}{K_s + S} + l_s\right) - k_{f2}MA + k_{r2}Cd_2 \tag{C.31}$$

$$\frac{dCd_1}{dt} = k_{f1}SA - k_{r1}Cd_1 - d_sCd_1 \tag{C.32}$$

$$\frac{dCd_2}{dt} = k_{f2}MA - k_{r2}Cd_2 - d_mCd_2 \tag{C.33}$$

$$\frac{dA}{dt} = -k_{f1}SA + k_{r1}Cd_1 + d_sCd_1 - k_{f2}MA + k_{r2}Cd_2 + d_mCd_2 \tag{C.34}$$

$$\frac{dG}{dt} = k_tM - d_pG \tag{C.35}$$

With a set amount of RNase in the pool, the overall degradation rate is faster at low induction level and slows down at high induction level. At low inducer concentration, fewer RNAs are transcribed. The free RNases can act on RNAs to form RNA-RNase complex which are then degraded. This allows for tight regulation of background signal leakage. As inducer concentration increases, more RNAs are produced and saturate the degradation machinery. As a result, there are more STAR molecules to activate the transcription of reporter and more mRNA being translated into GFP. This degradation mechanism elucidates how STAR mediated circuit is capable of reducing leak and improve dynamic range.

C.5 STAR Sequester impacts transfer function and dynamic range

In our efforts to further improve dynamic range in hybrid systems, we have found STAR sequesters as a prominent candidate at reducing off level signal leak. However, we have observed in experiments that hybrid system with sequester at on level have a greater signal deduction compared to off level. We thought this could also be explained by the degradation model: at off level, there are more free RNAses to degrade sequesters. At on level, the degradation machinery is saturated and more sequesters are produced than degraded. We have simulated the hybrid system with sequester to see whether the degradation model can reproduce experimental results. The tables below detail the model species and parameters for hybrid system with sequesters.

C.5.1 Model Species and Parameters

Table C.9: Species involved in STAR-hybrid system

Species	Description
S	STAR
Q	Sequester
$[SQ]$	STAR-Sequester Complex
M	mRNA of GFP
Cd_1	STAR-RNase Complex
Cd_2	Sequester-RNase Complex
Cd_3	STAR-Sequester-RNase Complex
Cd_4	mRNA-RNase Complex
A	Total RNase
G	GFP

Table C.10: Parameters involved in STAR-hybrid system

Parameter	number	Description	Value
X	1	Inducer Concentration	0.001 - 1e6
K_x	2	Activation coefficient of Inducer	20
l_x	3	Leak coefficient of Inducer	0.25
β_m	4	Transcriptional rate driven by constitutive promoter	4
d_s	5	Degradation rate of STAR	0.002
β_x	6	Max transcriptional rate driven by inducible promoter	2
K_s	7	Activation coefficient of STAR	2000
d_m	8	Degradation rate of mRNA	0.01
k_t	9	Translational rate of GFP	0.035
d_p	10	Degradation rate of GFP	0.0001
l_s	11	Leak coefficient of STAR $(0 - 1.0)$	0.001
k_{f1}	12	Forward binding constant of sRNA to RNase	1
k_{r1}	13	Reverse binding constant of sRNA to RNase	1000
k_{f2}	14	Forward binding constant of mRNA to RNase	1
k_{r2}	15	Reverse binding constant of mRNA to RNase	1000
β_Q	16	Transcriptional rate driven by constitutive promoter	1
k_b	17	Binding constant of STAR to Sequester	10

C.5.2 Equations

Similar to the degradation model proposed above, the dynamics of RNase activities on STAR, Sequester RNA, and mRNA are described by the following mechanisms:

$$S + A \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} Cd_1 \xrightarrow{d_s} A + \emptyset$$

$$Q + A \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} Cd_2 \xrightarrow{d_s} A + \emptyset$$

$$[S\,Q] + A \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} Cd_3 \xrightarrow{d_s} A + \emptyset$$

$$M + A \underset{k_{r2}}{\overset{k_{f2}}{\rightleftharpoons}} Cd_4 \xrightarrow{d_m} A + \emptyset$$

We assume that the forward (k_{f1}) and reverse (k_{r1}) binding constant of RNase to STAR, sequester, and STAR-sequester to be the same. Once the RNase-sRNA complexes (Cd_i) form, the sRNAs are degraded at the same speed d_s.

$$\begin{aligned}
\frac{dS}{dt} &= \beta_x\left(\frac{X}{K_x + X} + l_x\right) - k_{f1} S A + k_{r1} Cd_1 - k_b S Q && \text{(C.36)}\\
\frac{dQ}{dt} &= \beta_Q - k_{f1} Q A + k_{r1} Cd_2 - k_b S Q && \text{(C.37)}\\
\frac{d[S\,Q]}{dt} &= k_b S Q - k_{f1}[S\,Q] + k_{r1} Cd_3 && \text{(C.38)}\\
\frac{dM}{dt} &= \beta_m\left(\frac{S}{K_s + S} + l_s\right) - k_{f2} M A + k_{r2} Cd_2 && \text{(C.39)}\\
\frac{dCd_1}{dt} &= k_{f1} S A - k_{r1} Cd_1 - d_s Cd_1 && \text{(C.40)}\\
\frac{dCd_2}{dt} &= k_{f1} Q A - k_{r1} Cd_2 - d_s Cd_2 && \text{(C.41)}\\
\frac{dCd_3}{dt} &= k_{f1}[S\,Q] A - k_{r1} Cd_3 - d_s Cd_3 && \text{(C.42)}\\
\frac{dCd_4}{dt} &= k_{f2} M A - k_{r2} Cd_4 - d_m Cd_4 && \text{(C.43)}\\
\frac{dA}{dt} &= -k_{f1} S A + k_{r1} Cd_1 + d_s Cd_1 - k_{f1} Q A + k_{r1} Cd_2 + d_s Cd_2 \\
&\quad -k_{f1}[S\,Q] A + k_{r1} Cd_3 + d_s Cd_3 - k_{f2} M A + k_{r2} Cd_4 + d_m Cd_4 && \text{(C.44)}\\
\frac{dG}{dt} &= k_t M - d_p G && \text{(C.45)}
\end{aligned}$$

BIBLIOGRAPHY

[1] Uri Alon. *An Introduction to Systems Biology*. Design Principle of Biological Circuits. Chapman and Hall/CRC, 2006.

[2] Jens Bo Andersen, Claus Sternberg, Lars Kongsbak Poulsen, Sara Petersen Bjørn, Michael Givskov, and Søren Molin. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied and environmental microbiology*, 64(6):2240–2246, 1998.

[3] Ernesto Andrianantoandro, Subhayu Basu, David K Karig, and Ron Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, 2, May 2006.

[4] Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2(1):2006.0008–11, 2006.

[5] J E Bailey. Complex biology with no parameters. *Nature Biotechnology*, 19(6):503–504, June 2001.

[6] Arren Bar-Even, Johan Paulsson, Narendra Maheshri, Miri Carmi, Erin O'Shea, Yitzhak Pilpel, and Naama Barkai. Noise in protein expression scales with natural protein abundance. *Nature Genetics*, 38(6):636–643, May 2006.

[7] A Becskei and L Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, June 2000.

[8] Chase L Beisel and Christina D Smolke. Design principles for riboswitch function. *PLoS Computational Biology*, 5(4):e1000363, April 2009.

[9] David Bikard, David Bikard, Wenyan Jiang, Wenyan Jiang, Poulami Samai, Poulami Samai, Ann Hochschild, Ann Hochschild, Feng Zhang, Feng Zhang, Luciano A Marraffini, and Luciano A Marraffini. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research*, 41(15):7429–7437, August 2013.

[10] F R Blattner, G Plunkett, C A Bloch, N T Perna, V Burland, M Riley, J Collado-Vides, J D Glasner, C K Rode, G F Mayhew, J Gregor, N W Davis, H A Kirkpatrick, M A Goeden, D J Rose, B Mau, and Y Shao. The complete genome

sequence of Escherichia coli K-12. *Science*, 277(5331):1453–1462, September 1997.

[11] S Brantl and E G Wagner. Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181. *Molecular Microbiology*, 35(6):1469–1482, March 2000.

[12] S Brantl and E G H Wagner. An Antisense RNA-Mediated Transcriptional Attenuation Mechanism Functions in Escherichia coli. *Journal of Bacteriology*, 184(10):2740–2747, 2002.

[13] Jennifer A N Brophy and Christopher A Voigt. Principles of genetic circuit design. *Nature Publishing Group*, 11(5):508–520, May 2014.

[14] Kevin S Brown and James P Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 68(2 Pt 1):021904, August 2003.

[15] Barry Canton, Anna Labno, and Drew Endy. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology*, 26(7):787–793, July 2008.

[16] James M Carothers, Jonathan A Goler, Jonathan A Goler, Darmawi Juminaga, Darmawi Juminaga, and Jay D Keasling. Model-driven engineering of RNA devices to quantitatively program gene expression. *Science*, 334(6063):1716–1719, December 2011.

[17] T A Carrier and J D Keasling. Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli. *Biotechnology Progress*, 15(1):58–64, January 1999.

[18] Deepak Chandran, Frank T Bergmann, and Herbert M Sauro. Computer-aided design of biological circuits using TinkerCell. *Bioengineered bugs*, 1(4):274–281, July 2010.

[19] James Chappell, Kirsten Jensen, Kirsten Jensen, Paul S Freemont, and Paul S Freemont. Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Research*, 41(5):3471–3481, March 2013.

[20] James Chappell, Melissa K Takahashi, and Julius B Lucks. Creating small tran-

scription activating RNAs. *Nature Chemical Biology*, 11(3):214–220, March 2015.

[21] James Chappell, Melissa K Takahashi, Sarai Meyer, David Loughrey, Kyle E Watters, and Julius Lucks. The centrality of RNA for engineering gene expression. *Biotechnology Journal*, 8(12):1379–1395, December 2013.

[22] James Chappell, Alexandra Westbrook, Matthew Verosloff, and Julius B Lucks. Computational design of small transcription activating RNAs for versatile and dynamic gene regulation. *Nature Communications*, 8(1):1051, October 2017.

[23] James Chappell, Alexandra Westbrook, Matthew Verosloff, and Julius B Lucks. Computational design of small transcription activating rnas for versatile and dynamic gene regulation. *Nature communications*, 8(1):1051, 2017.

[24] Ying-Ja Chen, Peng Liu, Alec A K Nielsen, Jennifer A N Brophy, Kevin Clancy, Todd Peterson, and Christopher A Voigt. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature Publishing Group*, 10(7):659–664, July 2013.

[25] George M Church, Michael B Elowitz, Christina D Smolke, Christopher A Voigt, and Ron Weiss. Realizing the potential of synthetic biology. *Nature reviews. Molecular cell biology*, 15(4):289–294, April 2014.

[26] Kevin Clancy and Christopher A Voigt. Programming cells: towards an automated 'Genetic Compiler'. *Current opinion in biotechnology*, 21(4):572–581, August 2010.

[27] Chris D Cox, James M McCollum, Michael S Allen, Roy D Dar, and Michael L Simpson. Using noise to probe and characterize gene circuits. *Proceedings of the National Academy of Sciences*, 105(31):10809–10814, August 2008.

[28] K A Datsenko and B L Wanner. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences*, 97(12):6640–6645, June 2000.

[29] Domitilla Del Vecchio, Aaron J Dy, and Yili Qian. Control theory meets synthetic biology. *Journal of The Royal Society Interface*, 13(120):20160380–17, July 2016.

[30] M B Elowitz. Stochastic Gene Expression in a Single Cell. *Science*, 297(5584):1183–1186, 2002.

[31] M B Elowitz and S Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, January 2000.

[32] Carola Engler, Romy Kandzia, and Sylvestre Marillonnet. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*, 3(11):e3647, 2008.

[33] Michele Felletti, Julia Stifel, Lena A Wurmthaler, Sophie Geiger, and J ouml rg S Hartig. Twister ribozymes as highly versatile expression platforms for artificial riboswitches. *Nature Communications*, 7:1–8, September 2016.

[34] Ari E Friedland, Timothy K Lu, Xiao Wang, David Shi, George Church, and James J Collins. Synthetic gene networks that count. *Science*, 324(5931):1199–1202, May 2009.

[35] K G Gadkar, J Varner, and III Doyle FJ. Model identification of signal transduction networks from data using a state regulator problem. *Systems Biology, IEE Proceedings*, 2(1):17–30, March 2005.

[36] Jonathan Garamella, Ryan Marshall, Mark Rustad, and Vincent Noireaux. The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS synthetic biology*, 5(4):344–355, April 2016.

[37] T S Gardner, C R Cantor, and James J Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, January 2000.

[38] Luke A Gilbert, Matthew H Larson, Leonardo Morsut, Zairan Liu, Gloria A Brar, Sandra E Torres, Noam Stern-Ginossar, Onn Brandman, Evan H Whitehead, Jennifer A Doudna, Wendell A Lim, Jonathan S Weissman, and Lei S Qi. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2):442–451, July 2013.

[39] D Grate and C Wilson. Laser-mediated, site-specific inactivation of RNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America*, 96(11):6131–6136, May 1999.

[40] Alexander A Green, Pamela A Silver, James J Collins, and Peng Yin. Toehold switches: de-novo-designed regulators of gene expression. *Cell*, 159(4):925–939, November 2014.

[41] Călin C Guet, Luke Bruneaux, Taejin L Min, Dan Siegal-Gaskins, Israel Figueroa, Thierry Emonet, and Philippe Cluzel. Minimally invasive determi-

nation of mrna concentration in single living bacteria. *Nucleic acids research*, 36(12):e73–e73, 2008.

[42] J W Hearne. Sensitivity analysis of parameter combinations. *Applied Mathematical Modelling*, 9(2):106–108, 1985.

[43] Lisa M Hochrein, Maayan Schwarzkopf, Mona Shahgholi, Peng Yin, and Niles A Pierce. Conditional Dicer Substrate Formation via Shape and Sequence Transduction with Small Conditional RNAs. *Journal of the American Chemical Society*, 135(46):17322–17330, November 2013.

[44] C Eric Hodgman and Michael C Jewett. Cell-free synthetic biology: thinking outside the cell. *Metabolic Engineering*, 14(3):261–269, May 2012.

[45] Sara Hooshangi and Ron Weiss. The effect of negative feedback on noise propagation in transcriptional gene networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(2):026108–11, June 2006.

[46] Chelsea Y Hu, Melissa K Takahashi, Yan Zhang, and Julius B Lucks. Engineering a Functional Small RNA Negative Autoregulation Network with Model-Guided Design. *ACS synthetic biology*, 7(6):1507–1518, June 2018.

[47] Chelsea Y. Hu, Melissa K. Takahashi, Yan Zhang, and Julius B. Lucks. Engineering a functional small rna negative autoregulation network with model-guided design. *ACS Synthetic Biology*, 7(6):1507–1518, 2018. PMID: 29733627.

[48] Chelsea Y Hu, Jeffrey D Varner, and Julius B Lucks. Generating Effective Models and Parameters for RNA Genetic Circuits. *ACS synthetic biology*, 4(8):914–926, August 2015.

[49] Chelsea Y Hu, Jeffrey D Varner, and Julius B Lucks. Generating effective models and parameters for rna genetic circuits. *ACS synthetic biology*, 4(8):914–926, 2015.

[50] Razika Hussein, Razika Hussein, Han N Lim, and Han N Lim. Direct comparison of small RNA and transcription factor signaling. *Nucleic Acids Research*, 40(15):7269–7279, August 2012.

[51] T Ideker, V Thorsson, J A Ranish, R Christmas, J Buhler, J K Eng, R Bumgarner, D R Goodlett, R Aebersold, and L Hood. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292(5518):929–934, May 2001.

[52] Farren J Isaacs. Synthetic biology: Automated design of RNA devices. *Nature Chemical Biology*, 8(5):413–415, April 2012.

[53] Farren J Isaacs, Daniel J Dwyer, and James J Collins. RNA synthetic biology. *Nature Biotechnology*, 24(5):545–554, May 2006.

[54] Farren J Isaacs, Daniel J Dwyer, Chunming Ding, Dmitri D Pervouchine, Charles R Cantor, and James J Collins. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7):841–847, July 2004.

[55] F JACOB and J MONOD. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3):318–356, June 1961.

[56] Haeyoung Jeong, Young Mi Sim, Hyun Ju Kim, and Sang Jun Lee. Unveiling the Hybrid Genome Structure of Escherichia coli RR1 (HB101 RecA+). *Frontiers in microbiology*, 8:585, 2017.

[57] Eyal Karzbrun, Jonghyeon Shin, Roy H Bar-Ziv, and Vincent Noireaux. Coarse-Grained Dynamics of Protein Synthesis in a Cell-Free System. *Physical Review Letters*, 106(4):048104, January 2011.

[58] Yiannis N Kaznessis and Yiannis N Kaznessis. SynBioSS-aided design of synthetic biological constructs. *Methods in enzymology*, 498:137–152, 2011.

[59] Anastasia Khvorova, Aurélie Lescoute, Eric Westhof, and Sumedha D Jayasena. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nature Structural Biology*, 10(9):708–712, September 2003.

[60] Samira Kiani, Jacob Beal, Mohammad R Ebrahimkhani, Jin Huh, Richard N Hall, Zhen Xie, Yinqing Li, and Ron Weiss. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nature Publishing Group*, 11(7):723–726, July 2014.

[61] Hiroaki Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, March 2002.

[62] Anselm Levskaya, Aaron A Chevalier, Jeffrey J Tabor, Zachary Booth Simpson, Laura A Lavery, Matthew Levy, Eric A Davidson, Alexander Scouras, Andrew D Ellington, Edward M Marcotte, and Christopher A Voigt. Engineering Escherichia coli to see light. *Nature*, 438(7067):441–442, November 2005.

[63] David Loughrey, Kyle E Watters, Alexander H Settle, and Julius B Lucks. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Research*, 42(21):e165–e165, December 2014.

[64] Julius B Lucks, Lei Qi, Vivek K Mutalik, Denise Wang, and Adam P Arkin. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences*, 108(21):8617–8622, May 2011.

[65] B B Machta, B B Machta, R Chachra, R Chachra, M K Transtrum, M K Transtrum, J P Sethna, and J P Sethna. Parameter Space Compression Underlies Emergent Theories and Predictive Models. *Science*, 342(6158):604–607, October 2013.

[66] Daniel Madar, Erez Dekel, Anat Bren, and Uri Alon. Negative auto-regulation increases the input dynamic-range of the arabinose system of Escherichia coli. *BMC Systems Biology*, 5:111, 2011.

[67] Pankaj Mehta, Sidhartha Goyal, and Ned S Wingreen. A quantitative comparison of sRNA-based and protein-based gene regulation. *Molecular Systems Biology*, 4(1):221–10, 2008.

[68] Sarai Meyer, James Chappell, Sitara Sankar, Rebecca Chew, and Julius B Lucks. Improving fold activation of small transcription activating RNAs (STARs) with rational RNA engineering strategies. 2015.

[69] Neil Millar. Biology statistics made simple using Excel. *School Science Review*, pages 1–12, August 2002.

[70] R Milo, S Shen-Orr, S Itzkovitz, N Kashtan, D Chklovskii, and U Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, October 2002.

[71] Alexander Mitsos, Ioannis N Melas, Melody K Morris, Julio Saez-Rodriguez, Douglas A Lauffenburger, and Leonidas G Alexopoulos. Non Linear Programming (NLP) formulation for quantitative modeling of protein signal transduction pathways. *PLoS ONE*, 7(11):e50085, 2012.

[72] J Monod, J Wyman, and J P Changeux. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, 12:88–118, May 1965.

[73] Jacques Monod and François Jacob. General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harbor Symposia on Quantitative Biology*, 26(0):389–401, January 1961.

[74] Tae Seok Moon, Chunbo Lou, Alvin Tamsir, Brynne C Stanton, and Christopher A Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491(7423):249–253, October 2012.

[75] Dmitry Nevozhay, Rhys M Adams, Kevin F Murphy, Kresimir Josic, and Gábor Balázsi. Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proceedings of the National Academy of Sciences*, 106(13):5123–5128, March 2009.

[76] Alec A K Nielsen, Thomas H Segall-Shapiro, and Christopher A Voigt. Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. *Current Opinion in Chemical Biology*, 17(6):878–892, December 2013.

[77] Alec A K Nielsen and Christopher A Voigt. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Molecular Systems Biology*, 10(11):763, November 2014.

[78] Lior Nissim, Samuel D Perli, Alexandra Fridkin, Pablo Perez-Pinera, and Timothy K Lu. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular Cell*, 54(4):698–710, May 2014.

[79] Vincent Noireaux, Roy Bar-Ziv, Roy Bar-Ziv, Albert Libchaber, and Albert Libchaber. Principles of cell-free genetic circuit assembly. *Proceedings of the National Academy of Sciences*, 100(22):12672–12677, October 2003.

[80] Bela Novak and John J Tyson. Design principles of biochemical oscillators. *Nature reviews. Molecular cell biology*, 9(12):981–991, December 2008.

[81] R P Novick, R P Novick, S Iordanescu, S Iordanescu, S J Projan, S J Projan, J Kornblum, J Kornblum, I Edelman, and I Edelman. Pt181 Plasmid Replication Is Regulated by a Countertranscript-Driven Transcriptional Attenuator. *Cell*, 59(2):395–404, 1989.

[82] Erin L O'Brien, Elizabeth Van Itallie, and Matthew R Bennett. Modeling synthetic gene oscillators. *Mathematical biosciences*, 236(1):1–15, March 2012.

[83] Evan J Olson, Lucas A Hartsough, Brian P Landry, Raghav Shroff, and Jeffrey J Tabor. Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nature Methods*, 11(4):449–455, April 2014.

[84] Keith Pardee, Alexander A Green, Tom Ferrante, D Ewen Cameron, Ajay DaleyKeyser, Peng Yin, and James J Collins. Paper-based synthetic gene networks. *Cell*, 159(4):940–954, November 2014.

[85] Keith Pardee, Alexander A Green, Melissa K Takahashi, Dana Braff, Guillaume Lambert, Jeong Wook Lee, Tom Ferrante, Duo Ma, Nina Donghia, Melina Fan, Nichole M Daringer, Irene Bosch, Dawn M Dudley, David H O'Connor, Lee Gehrke, and James J Collins. Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell*, 165(5):1255–1266, May 2016.

[86] Jean-Denis Pédelacq, Jean-Denis Pédelacq, Stéphanie Cabantous, Stéphanie Cabantous, Timothy Tran, Timothy Tran, Thomas C Terwilliger, Thomas C Terwilliger, Geoffrey S Waldo, and Geoffrey S Waldo. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, 24(1):79–88, January 2006.

[87] Priscilla E M Purnick and Ron Weiss. The second wave of synthetic biology: from modules to systems. *Nature reviews. Molecular cell biology*, 10(6):410–422, June 2009.

[88] Lei S Qi and Adam P Arkin. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nature Reviews Microbiology*, 12(5):341–354, May 2014.

[89] Lei S Qi, Matthew H Larson, Luke A Gilbert, Jennifer A Doudna, Jonathan S Weissman, Adam P Arkin, and Wendell A Lim. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183, February 2013.

[90] Nitzan Rosenfeld and Uri Alon. Response delays and the structure of transcription networks. *Journal of Molecular Biology*, 329(4):645–654, June 2003.

[91] Nitzan Rosenfeld, Michael B Elowitz, and Uri Alon. Negative Autoregulation Speeds the Response Times of Transcription Networks. *Journal of Molecular Biology*, 323(5):785–793, November 2002.

[92] Marc R Roussel, Marc R Roussel, Rui Zhu, and Rui Zhu. Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Physical Biology*, 3(4):274–284, November 2006.

[93] M A Savageau. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature*, 252(5484):546–549, December 1974.

[94] Piali Sengupta and Paul Garrity. Sensing temperature. *Current Biology*, 23(8):R304–R307, April 2013.

[95] Alexander Serganov and Evgeny Nudler. A decade of riboswitches. *Cell*, 152(1-2):17–24, January 2013.

[96] Nathan C Shaner, Paul A Steinbach, and Roger Y Tsien. A guide to choosing fluorescent proteins. *Nature Methods*, 2(12):905–909, December 2005.

[97] Vinay Shimoga, Jacob T White, Yi Li, Eduardo Sontag, and Leonidas Bleris. Synthetic mammalian transgene negative autoregulation. *Molecular Systems Biology*, 9:1–7, June 2013.

[98] Jonghyeon Shin and Vincent Noireaux. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *Journal of biological engineering*, 4:8–8, January 2010.

[99] Michael L Simpson, Chris D Cox, and Gary S Sayler. Frequency domain analysis of noise in autoregulated gene circuits. *Proceedings of the National Academy of Sciences*, 100(8):4551–4556, April 2003.

[100] Abhyudai Singh and Joao P Hespanha. Optimal Feedback Strength for Noise Suppression in Autoregulatory Gene Networks. *Biophysj*, 96(10):4013–4023, May 2009.

[101] Dov J Stekel and Dafyd J Jenkins. Strong negative self regulation of prokaryotic transcription factors increases the intrinsic noise of protein expression. *BMC Systems Biology*, 2:6, January 2008.

[102] Jason T Stevens, Jason T Stevens, James M Carothers, and James M Carothers. Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways. *ACS synthetic biology*, 4(2):107–115, February 2015.

[103] Ann M Stock, Victoria L Robinson, and Paul N Goudreau. Two-Component Signal Transduction. *Annual Review of Biochemistry*, 69(1):183–215, 2000.

[104] Jesse Stricker, Scott Cookson, Matthew R Bennett, William H Mather, Lev S

Tsimring, and Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, November 2008.

[105] Eric J Strobel, Kyle E Watters, Yuri Nedialkov, Irina Artsimovitch, and Julius B Lucks. Distributed biotin-streptavidin transcription roadblocks for mapping cotranscriptional RNA folding. *Nucleic Acids Research*, 45(12):e109, July 2017.

[106] Zachary Z Sun, Clarmyra A Hayes, Jonghyeon Shin, Filippo Caschera, Richard M Murray, and Vincent Noireaux. Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *Journal of visualized experiments : JoVE*, (79):e50762–e50762, September 2013.

[107] Jeffrey J Tabor, Howard M Salis, Zachary Booth Simpson, Aaron A Chevalier, Anselm Levskaya, Edward M Marcotte, Christopher A Voigt, and Andrew D Ellington. A Synthetic Genetic Edge Detection Program. *Cell*, 137(7):1272–1281, June 2009.

[108] Melissa K Takahashi, James Chappell, Clarmyra A Hayes, Zachary Z Sun, Jongmin Kim, Jongmin Kim, Vipul Singhal, Vipul Singhal, Kevin J Spring, Kevin J Spring, Shaima Al-Khabouri, Shaima Al-Khabouri, Christopher P Fall, Christopher P Fall, Vincent Noireaux, Richard M Murray, and Julius B Lucks. Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems. *ACS synthetic biology*, 4(5):503–515, May 2015.

[109] Melissa K Takahashi, Clarmyra A Hayes, James Chappell, Zachary Z Sun, Richard M Murray, Vincent Noireaux, and Julius B Lucks. Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. *Methods (San Diego, Calif.)*, 86:60–72, September 2015.

[110] Melissa K Takahashi and Julius B Lucks. A modular strategy for engineering orthogonal chimeric RNA transcription regulators. *Nucleic Acids Research*, 41(15):7577–7588, August 2013.

[111] Melissa K Takahashi, Kyle E Watters, Paul M Gasper, Timothy R Abbott, Paul D Carlson, Alan A Chen, and Julius B Lucks. Using in-cell SHAPE-Seq and simulations to probe structure-function design principles of RNA transcriptional regulators. *RNA*, 22(6):920–933, June 2016.

[112] Alvin Tamsir, Jeffrey J Tabor, and Christopher A Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature*, 469(7329):212–215, December 2010.

[113] M Thattai and A van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, July 2001.

[114] J J Tyson, K Chen, and B Novak. Network dynamics and cell physiology. *Nature reviews. Molecular cell biology*, 2(12):908–916, December 2001.

[115] Neelam Verma and Minni Singh. Biosensors for heavy metals. *BioMetals*, 18(2):121–129, April 2005.

[116] U Vogel and K F Jensen. The RNA chain elongation rate in Escherichia coli depends on the growth rate. *Journal of Bacteriology*, 176(10):2807–2813, May 1994.

[117] Christopher A Voigt. Genetic parts to program bacteria. *Current opinion in biotechnology*, 17(5):548–557, October 2006.

[118] Margaritis Voliotis and Clive G Bowsher. The magnitude and colour of noise in genetic negative feedback systems. *Nucleic Acids Research*, 40(15):7084–7095, May 2012.

[119] Justin D Vrana, William E Voje, James M Carothers, Eric Klavins, and Miles W Gander. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nature Communications*, 8:1–11, May 2017.

[120] Manja Wachsmuth, Sven Findeiß, Nadine Weissheimer, Peter F Stadler, and Mario Mörl. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Research*, 41(4):2541–2551, February 2013.

[121] Yen-Hsiang Wang, Maureen McKeague, Tammy M Hsu, and Christina D Smolke. Design and Construction of Generalizable RNA-Protein Hybrid Controllers by Level-Matched Genetic Signal Amplification. *Cell Systems*, 3(6):549–562.e7, December 2016.

[122] Kyle E Watters, Eric J Strobel, Angela M Yu, John T Lis, and Julius B Lucks. Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nature Publishing Group*, 23(12):1124–1131, December 2016.

[123] Eric W Weisstein. Least Squares Fitting. 2002.

[124] Alexandra M Westbrook and Julius B Lucks. Achieving large dynamic range control of gene expression with a compact RNA transcription–translation regulator. *Nucleic Acids Research*, 45(9):5614–5624, April 2017.

[125] Kristoffer Skovbo Winther, Mohammad Roghanian, and Kenn Gerdes. Activation of the Stringent Response by Loading of RelA-tRNA Complexes at the Ribosomal A-Site. *Molecular Cell*, 70(1):95–105.e4, April 2018.

[126] Edward Yang, Edward Yang, Erik van Nimwegen, Erik van Nimwegen, Mihaela Zavolan, Mihaela Zavolan, Nikolaus Rajewsky, Nikolaus Rajewsky, Mark Schroeder, Mark Schroeder, Marcelo Magnasco, Marcelo Magnasco, James E Darnell, and James E Darnell. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Research*, 13(8):1863–1872, August 2003.

[127] K Zhen Yao, Benjamin M Shaw, Bo Kou, Kim B McAuley, and D W Bacon. Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design. *Polymer Reaction Engineering*, 11(3):563–588, 2003.

[128] K Zhen Yao, Benjamin M Shaw, Bo Kou, Kim B McAuley, and DW Bacon. Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimability and experimental design. *Polymer Reaction Engineering*, 11(3):563–588, 2003.

[129] Ji Yu, Ji Yu, Jie Xiao, Jie Xiao, Xiaojia Ren, Xiaojia Ren, Kaiqin Lao, Kaiqin Lao, X Sunney Xie, and X Sunney Xie. Probing gene expression in live cells, one protein molecule at a time. *Science*, 311(5767):1600–1603, March 2006.

[130] Joseph N Zadeh, Conrad D Steenberg, Justin S Bois, Brian R Wolfe, Marshall B Pierce, Asif R Khan, Robert M Dirks, and Niles A Pierce. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32(1):170–173, January 2011.

[131] Alon Zaslaver, Avi E Mayo, Revital Rosenberg, Pnina Bashkin, Hila Sberro, Miri Tsalyuk, Michael G Surette, and Uri Alon. Just-in-time transcription program in metabolic pathways. *Nature Genetics*, 36(5):486–491, May 2004.

[132] Yonglong Zhang, Junjie Zhang, Klaus P Hoeflich, Mitsuhiko Ikura, Guoliang Qing, and Masayori Inouye. MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. *Molecular Cell*, 12(4):913–923, October 2003.